Progress in
EXPERIMENTAL PERSONALITY RESEARCH

VOLUME 1

CONTRIBUTORS TO THIS VOLUME

DONN BYRNE

JEAN P. CHAPMAN

LOREN J. CHAPMAN

VICTOR B. CLINE

DOUGLAS T. KENNY

S. H. LOVIBOND

GLENN A. MILLER

ROBERT ROSENTHAL

P. H. VENABLES

PROGRESS IN

Experimental Personality Research

Edited by Brendan A. Maher

HARVARD UNIVERSITY CENTER
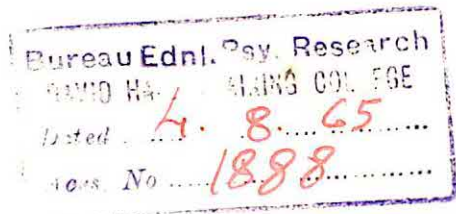FOR RESEARCH IN PERSONALITY
CAMBRIDGE, MASSACHUSETTS

VOLUME 1

1964

ACADEMIC PRESS New York and London

CONTRIBUTORS

Number in parentheses indicates the page on which the author's contribution begins.

DONN BYRNE (169), *Department of Psychology, The University of Texas, Austin, Texas*

JEAN P. CHAPMAN (49), *Department of Psychology, Southern Illinois University, Carbondale, Illinois*

LOREN J. CHAPMAN (49), *Department of Psychology, Southern Illinois University, Carbondale, Illinois*

VICTOR B. CLINE (221), *University of Utah, Salt Lake City, Utah*

DOUGLAS T. KENNY (285), *Harvard University, Cambridge, Massachusetts*

S. H. LOVIBOND[1] (115), *Department of Psychology, University of Adelaide, Adelaide, South Australia*

GLENN A. MILLER (49), *Department of Psychology, Southern Illinois University, Carbondale, Illinois*

ROBERT ROSENTHAL (79), *Harvard University, Cambridge, Massachusetts*

P. H. VENABLES (1), *Medical Research Council, Social Psychiatry Research Unit, Institute of Psychiatry, Maudsley Hospital, London, England*

---

[1] Present address: The Psychology Department, Dalhousie University, Halifax, Nova Scotia, Canada.

v

PREFACE

Personality psychology has changed during the past two or three decades. The change has been marked by a move away from theorizing on the grand scale and toward a greater concern with obtaining empirical answers to questions of manageable and modest proportions. By the same token, research in personality has turned increasingly to the methods and concepts of other areas of behavioral science; the shift from the study of the single case to the use of the controlled experiment is perhaps the most striking instance of this.

With this movement, the lines of demarcation between personality psychology and other areas of psychology have become blurred. The student of personality is faced with considerable extension of his field of interest, and the usual concomitant difficulty in remaining *au fait* with the progress that is made in important technical areas.

In this series, contributors will present both a summary of present knowledge on a specialized topic together with original data from their own investigations of the problem. This book is intended for psychologists and students of psychology, psychopathologists and all whose work requires knowledge of recent developments in the study of personality.

Two original and systematic contributions to the understanding of schizophrenia appear in this first volume. Other papers deal with interpersonal perception, experimenter bias, the personality correlates of repression and sensitization, conditioning and personality, and processes in projective perception.

Much editorial help has been given in the preparation of this volume by Kathleen Sylva.

BRENDAN A. MAHER

Cambridge, Massachusetts
August 1964

CONTENTS

Input Dysfunction in Schizophrenia

P. H. Venables

A Theory of Verbal Behavior in Schizophrenia

Loren J. Chapman, Jean P. Chapman, and
Glenn A. Miller

The Effect of the Experimenter on the Results of Psychological Research

Robert Rosenthal

INPUT DYSFUNCTION IN SCHIZOPHRENIA

P. H. Venables

MEDICAL RESEARCH COUNCIL, SOCIAL PSYCHIATRY RESEARCH UNIT,
INSTITUTE OF PSYCHIATRY, MAUDSLEY HOSPITAL, LONDON, ENGLAND

I. Introduction

A. SCOPE OF THE REVIEW

"Input dysfunction," rather than "disorder of perception or sensation," has been chosen as the title of this review because it was felt necessary to avoid the restriction of emphasis which the older terminology connotes. Implicitly, if not explicitly, sensation and perception as elements in cognitive processes have been thought of as confined to the functions of the direct sensory pathways and the receiving and association areas of the cortex. It is necessary at the present stage of our knowledge to bear in mind the many processes which interact before an item of sensory input can arise in consciousness. Not only does sensory input reach the cortex by the classical sensory pathways but it influences mesencephalic and diencephalic structures, via collaterals from the direct pathways to modify states of awareness, attention, arousal, and emotion. In the case of noninformative input such as noise, the main result of this input may in fact be its action upon subcortical structures. In turn the state of the subcortical structures may, by their effects at peripheral re-

1

ceptor organs and at all levels up to the cortex, modify the pattern of sensory input. Neurophysiological studies also suggest that the cortex can have a profound effect on pathways through the brain stem and "indicate the critical role of the cortex, with its highly discriminative properties in the selection and transmission of sensory input" (Samuels, 1959).

On a psychological level, owing possibly to a certain amount of re-thinking which has occurred following the incorporation of information theory and related concepts into everyday psychological currency, the old idea that we cannot consider a sensory impression to be an isolated event has received additional emphasis. An item of sensory input must be considered as one in a set of possible input items having different prob-abilities of occurrence. It must be integrated into a pattern with respect to past events held in memory, it must indeed be *apperceived* to use an old-fashioned but extremely useful term. Thus the relative importance of different items, the degree to which they appear to the subject to form a subset of mutual relevance out of all other items in the sensory field, becomes a matter of necessary consideration. This leads our think-ing back to such a concept as the span of attention, or to the similar but not identical idea, range of cue utilization, defined by Easterbrook (1959) as the "total number of environmental cues in any situation that an organism observes, maintains an orientation towards, responds to or as-sociates with a response." From what has been said earlier, it is apparent that degree of attention is under cortical and subcortical control and is thus subject not only to the temporary phasic emotional states of the organism, but also to the longer-term cortical-subcortical disarrangements that are manifest in mental illness.

Because sensory input must always be thought of as having impor-tance to the organism insofar as it is related to, and may be integrated with, its past experience, the distinction among sensation, perception, and thinking becomes blurred. This means that in dealing with experi-mental studies which are supposed to be concerned with input dysfunc-tion in schizophrenia, the vast literature on thought disorder must always loom large. The boundary between the terms of reference of this review and another covering thought disorder would be vague and difficult to define. Nevertheless the attempt will be made to confine attention to studies which are devoted for the most part to consideration of input into the organism, and only where relevant to the matter of how it is dealt with. The reader's indulgence must, however, be sought in the face of such clearly justifiable statements as the following:

"Norman Cameron has defined overinclusive thinking as the in-ability to preserve conceptual boundaries which results in the incorpora-tion of irrelevant ideas, making thinking more abstract and less lucid.

Payne *et al.* (1959) have extended the definition, regarding it as an attention defect; as some breakdown in a hypothetical filter mechanism which normally screened out those stimuli both internal and external which are irrelevant to a task in hand, to allow the most efficient processing of incoming information" (Payne, 1962).

It is to a large extent with the subject of Payne's extended definition that this review will be concerned.

Work on disturbances of sensation and perception in schizophrenia earlier in the century received limited attention, perhaps because of the findings of Kraepelin, who, using Wundtian techniques, reported little sensory disturbance in schizophrenia. This position was supported by the later pronouncements of Bleuler on the "clear sensorium" of schizophrenics. It is perhaps to the Gestalt and psychoanalytic schools that the present work on input dysfunction in schizophrenia may most clearly be traced; and it is in these historical sources that the origins of the curious patchiness of coverage may be found. As examples we find considerable literature on size-constancy abnormalities, with an extension into work on the perceived size of body parts, and little or no work on other constancy phenomena. There is quite a large amount of work on the effect of social and aversive stimuli on conditioning and performance of schizophrenics and a comparative dearth of studies on the effect of a change in the parameters of simpler sensory stimuli.

Alongside, and perhaps as a consequence of this patchwork coverage of possible dysfunctions in schizophrenia, there is an incompleteness of reporting manifest in many studies that makes difficulties for the later worker who needs more detailed information to compare with his own results. This is particularly the case in reporting the nature of subject groups. This may be due to the bias of the experimenter, who, perhaps testing an idea that all psychotics show a certain abnormality, lumps together schizophrenics and depressives whether chronic or acute.

It is with some relief that a tendency for more complete reporting may be noticed in more recent papers. This, of course, is a cumulative phenomenon and is increasingly reinforced as experimenters report differences between, for instance, paranoid and nonparanoid subjects, between chronic and acute patients, or between process or reactive subgroups. While this is a step in the right direction the warnings of Shakow (1963) should be heeded. Dichotomies themselves lead to oversimplification and frequently overlap. A rating scale to measure withdrawal may possibly measure nearly the same thing as deterioration or elements of process schizophrenia; the label paranoid when based on a scale of nurses' ratings may include a mixture of incoherent and paranoid tendencies.

Equally open to the possibilities of misinterpretation is the acute-chronic dichotomy. A dividing line of more or less than 2 years stay in hospital may be a perfectly reasonable cutting point when based on the probabilities of discharge from hospital (Brown, 1960). However, the transition from the acute to the chronic phase of the illness is part of the process of the disease involving an interaction between the patient and his environment and the determination of the boundary between the two states may be much less easy.

B. Some Theoretical Viewpoints

The distinction between the acute and chronic phases of the illness is fundamental to the interpretation of much of the literature which will be reviewed. Unfortunately the evidence both on the existence of a clear distinction between these two states and a difference in underlying mechanisms, should such a distinction exist, is incomplete. Two types of theory will be considered. One, typified, for instance, by Mednick (1958) and Fish (1961), suggests a state of high drive or arousal in the acute phase which subsides in the chronic phase. The other type of theory, suggested by, for instance, Hoffer and Osmond (1955) and Weckowicz (1958), is of a parasympathetic imbalance in the acute phase and a late change to sympathetic imbalance as the patient enters the chronic phase. Both types of theory draw much the same sort of distinction between the clinical pictures in each phase, but explain them rather differently.

The point of view of the latter type of theory is stated clearly by Weckowicz (1958). "Early in the illness, function shifts in the direction of parasympathetic activity and through negative feedback mechanisms [see the work of Callaway, Section II,A] size constancy increases. This signifies increased constancy[1] of perception which in its turn may be interpreted as increased awareness of objects in the environment. The patients become sensitive, suspicious, develop ideas of reference, attach important meanings to insignificant events and eventually develop delusions. They react to the environment with increased anxiety, because the low activity of the diencephalic sympathetic centres is accompanied by high reactivity. However, as the disease progresses the function of these centres gradually shifts in the opposite direction, namely towards sympathetic activity." This, Weckowicz says, is accompanied by decreased size constancy, which is indicative of a diminished awareness of objects in the environment. Because of the high level of

---

[1] Weckowicz is here discussing the results of an experiment on size constancy and the term "increased constancy," or alternatively "overconstancy," implies that the size of a distant object is overestimated.

sympathetic activity, reactivity is low and the patients display "flatness of affect"; they also become withdrawn, "autistic," lose interest in their environment, and sometimes have emotional outbursts which often cannot be related to external events. This suggestion of early parasympathetic and later sympathetic imbalance is in accord with the viewpoint expressed by Hoffer and Osmond (1955), derived on the basis of biochemical theory.

The increased awareness of the environment which is reported here as being characteristic of the acute phase is graphically documented by McGhie and Chapman (1961). Patients are swamped by a flood of sensory input which they are unable to control. They report:

"Everything seems to grip my attention although I am not particularly interested in anything. I am speaking to you just now but I can hear noises going on next door and in the corridor. I find it difficult to shut these out and it makes it more difficult to concentrate on what I am saying to you."

"Things are coming in too fast. I lose my grip of it and get lost. I am attending to everything at once and as a result I do not really attend to anything."

Not only are there these complaints of inability to focus attention, to select between items of sensory input, but there are impressions of abnormalities in the quality of the sensory input.

"Noises seem to be louder to me than they were before . . . . I notice it most with background noises."

"Colours seem to be brighter now almost as if they are luminous."

"The colours of things seem to be much clearer and yet at the same time there is something missing. The things I look at seem to be flatter as if you were looking just at the surface."

Here we have reports of the lack of distinction between relevant and irrelevant background material and of evident changes in distance constancy.

The lack of ability to distinguish "figure" from "ground" is particularly apparent in the perception of speech.

"When people are talking I just get scraps of it. If it is just one person who is speaking that's not so bad but if the others join in then I can't pick it up at all. I just can't get into time with that conversation. It makes me feel open as if things are closing in on me and I have lost control."

"I'm a good listener but often I'm not really taking it in. I nod my head and smile but it's just a lot of jumbled up words to me."

In addition to this disturbance of ability to distinguish relevant from irrelevant items and to place them in their correct order, there also

appears to be an inability to absorb items of information if they are presented too quickly. This is particularly the case where an increase in the rate of sensory input is brought about by the patient's own movement.

"When I move quickly it's a strain on me. Things go too quick for my mind. They get blurred and it's like being blind. It's as if you were seeing a picture one moment and another picture the next. I just stop and watch my feet. Everything is all right if I stop, but if I start moving again I lose control."

Heightened awareness of bodily sensations brings about a loss of automatic spontaneity of action; no longer can a sequence of motor acts be carried out without conscious deliberation because the patient is made aware of each stage. Thus, as the authors say, "each action now has to be planned and executed step by step with a great deal of conscious deliberation. The patient finds himself becoming increasingly 'self conscious' in an entirely literal sense."

"People just do things but I have to watch first to see how you do things. I have to think out most things first and know how to do them before I do them."

This clinical material has been quoted in some detail because it presents in a particularly graphic manner the type of dysfunction that is manifest in a large number of patients in the acute phase and which has to be borne in mind when considering experimental studies.

Because such insightful reporting of the nature of sensation in the chronic phase of the illness does not seem to be available, the main evidence for a decreased awareness of the full range of the subject's environment must rely to a large extent on experimental data, which will be presented later (Section II,B).

McGhie and Chapman present suggestions to account for the clinical material which they describe. Their present theoretical position is perhaps all the more interesting because, starting out from an investigation which was psychoanalytic in orientation, they are led by their observations to reject the proposal (Federn, 1953) that early schizophrenic symptoms are defensive activities related to unconscious conflicts and to suggest that the basic pathological breakdown is particularly to be found in the process of perception. From this they say that "if schizophrenia is a disease which has its basic effect in a disruption of the control of attention, then the reticular system may be the main pathological site."

A view which is in many ways similar, but has a somewhat different emphasis, is that of Fish (1961). Starting from a Gestalt analysis of the stages of schizophrenic disorganization advocated by Conrad (1958), which he modifies in terms of Hebb's (1949) theory of perceptual

organization, Fish reaches the conclusion that early phases of acute schizophrenia are due to "overactivity of the reticular system producing an undue diversion and disruption of central processes by the sensory input." In the presence of reticular overactivity sensory events thus acquire greater than normal significance, thought processes are disorganized, and anxiety results. When this disruption of the central processes by sensory input occurs frequently a "feeling of some unknown significance is produced by the increased impressiveness of all perceptions." From these processes delusions arise.

As the next stage, Fish proposes that parallel central reverberatory processes develop which are detached from sensory input and are not reorganized by it. This state of affairs may continue to exist in the chronic state after the high reticular activity which engendered it has ceased. It is of interest to note that Fish proposes three types of chronic state: In one the reticular activity does not subside and a paranoid illness remains in which delusions continue to be affect laden. Second and third, reticular activity subsides but the cortical reorganization remains either in the same form as in the acute processes, or in a form different from this but less efficient than the premorbid state. Fish backs up his case for reticular overactivity by several pointers from physiology and pharmacology. For instance, noting that amphetamine overdosage will produce a psychosis which is clinically undistinguishable from paranoid schizophrenia, he points out that amphetamine produces an arousal response in the EEG in animals which is similar to that produced by reticular stimulation. This psychotomimetic effect of amphetamine should be borne in mind when experiments on the effect of alerting agents on the breadth of attention are discussed at a later stage (Section II,A).

In contrast to the previous theories, in that it remains on a behavioral level, is the "learning theory approach" of Mednick (1958). In spite of this difference the picture which results has similarities to many of the foregoing theories and also to that of Rosenzweig (1955). In both these proposals it is the erection of unrealistic defences against anxiety which results in psychotic behavior. In Rosenzweig's theory the anxiety results because of disturbances in central processes which normally integrate ideational and affective processes; in Mednick's theory the anxiety results from a combination of factors due to a high drive state which is said to characterize the acute stage of schizophrenia. This high drive has three results: "(a) schizophrenics more easily acquire a conditioned response; (b) schizophrenics show greater stimulus generalization responsiveness; (c) schizophrenics have great difficulty performing well in complex situations being plagued by irrelevant tangential

associative responses competing with adequate modes of response."
Mednick goes on to elaborate his theory in terms of a reciprocal augmen-
tation mechanism. An untoward event will, in a person with an already
high drive level, raise his anxiety level; consequent upon this, the range
of stimulus generalization will increase and further stimuli will acquire
anxiety-producing qualities. This process continues until the behavior of
the individual becomes noticeably unusual. High drive level keeps
thoughts racing in the patient's mind and a rationalizing solution is
evoked to reduce anxiety. Lack of discrimination is brought about by
high generalization and any associated response, however tangential,
may produce anxiety reduction resulting in thought disorder. The
anxiety reduction which may be achieved by merely thinking irrelevant
thoughts results in the patient displaying less and less emotionality and
the chronic stage of the illness is reached.

These theories which have been presented describe rather similar
clinical pictures, the acute stage with its excesss of sensory input, and
overgeneralization followed by a transition to a chronic stage with
diminished awareness of outside stimuli and withdrawal. Although the
description of the clinical state roughly corresponds in the various
theories, the suggested underlying mechanisms tend to be somewhat
opposed. On the one hand Weckowicz, and Hoffer and Osmond, suggest
a parasympathetic overactivity in the acute stage followed by a transition
to a sympathetically aroused and cortically activated picture in the chronic
patient. On the other hand Fish and Mednick propose high reticular
activity or high drive in the acute state which later subsides as the patient
becomes chronic. The remainder of this chapter will be concerned not
so much with theories but with an examination of the experimental
justification for them, and an attempt to see how far the rather divergent
viewpoints about mechanisms may be either reconciled or verified.

II. The Span of Attention

A. Conditions Bringing about Alterations in Span of Attention in Normal Subjects

It was suggested earlier (Section I,A) that perception and attention
were influenced by subcortical as well as cortical mechanisms. An
examination of relevant experiments on normal subjects under condi-
tions which might be expected to have subcortical effects is valuable
insofar as it may enable insight to be gained which will be helpful in
the interpretation of experiments with schizophrenic subjects. Perhaps
the most relevant of these are a series of studies by Callaway and his
associates which are concerned with the effect of certain drugs and

procedures on perceptual tasks which are interpretable in terms of extent of attention.

Callaway and Thompson (1953) proposed an hypothetical mechanism which they then proceeded to test, by which the activity of the sympathetic nervous system and the breadth of attention are related. They suggested that a threat to the organism which induces a sympathetic discharge must in turn bring about a rise in the threshold of the input system which will tend to reduce the perception of the threat and hence reduce the sympathetic discharge (i.e., a negative feedback mechanism). Otherwise if increased sympathetic discharge induced a lowering of threshold there would be an increasing rise of sympathetic discharge to an intolerable limit. (This model was later modified because it was found that the common factor in treatments producing narrowing of attention was that they all produced an alert EEG.) The model was tested by the use of a size-constancy experiment. They suggested: "We should expect increased sympathetic activity to decrease that general awareness of gradients by which we correct our shrunken retinal image of distant objects. Objects would then be perceived as smaller than their true size and would seem to approach a size more nearly proportional to their retinal image size."

The size-constancy experiment which was carried out involved the adjustment of the size of a projected rectangular patch of light 200 cm away from the subject to match the size of a rectangular card held in the hand. To induce sympathetic discharge a cold pressor test was employed in which the subject's foot was immersed in a bucket of ice water. This procedure evoked a reduction of size constancy in comparison to that obtained under nonstimulating conditions. Because the cold pressor procedure results in hypertension which would produce increased retinal blood flow in addition to sympathetic discharge, the experiment was repeated with inhalation of amyl nitrite as the sympathetic stimulant. In contrast to the previous procedure amyl nitrite induces hypotension. The previous result was replicated. Various possibilities which might result in the perceptual effect were considered in further experiments; explanations in terms of change of perceived distance to the adjustable stimulus and change of actual size of retinal image akin to atropine micropsia were tested and dismissed. An experimental situation in which there was a reduction in subsidiary cues abolished the effect earlier noticed so that the conclusion was reached that decreased size constancy after induced sympathetic adrenergic discharge resulted from reduced awareness of subsidiary cues which normally act to correct the diminution in retinal size of distant objects. Callaway and Dembo (1958) and Callaway (1959) reported further work on the phenomenon of changes in

breadth of attention which tended to confirm the earlier findings. The findings reported with amyl nitrite were replicated with a further group of subjects and other experiments using epinephrine and methamphetamine showed similar changes in size judgment. Size-constancy experiments were eventually abandoned because of the difficulty in showing directly whether the effect was due to decreased awareness of subsidiary cues or had some other explanation. Among others a task which these workers used for further experimentation was that of perceiving the statistical structure of the pattern of onset of lights in a binary guessing game. It has been shown (e.g., Hake and Hyman, 1953) that subjects' guesses on a binary choice game such as this tend to approach very closely the actual frequency with which the two choices are reinforced as correct. Callaway et al. argue that this procedure may be used as a measure of narrowed attention. "In the guessing game the central focus of attention will be held alternately by the current guess and the current answer. . . . Since past answers are continually peripheral to current guesses, guess frequency response to answer frequency change reflects the influence of peripheral factors upon behaviour. In other words it can be used to indicate narrowed attention." In specific terms the experimenters predicted that methamphetamine would produce narrowing of attention and hence would result in subjects responding more slowly than control subjects to a change in the imposed statistical pattern.

It should perhaps be noted at this point that although what apparently are on the periphery here are memory traces and not true sensory input, we have a position which is analogous to an information-conveying input situation where the immediately present stimulus should be looked at as one of a set of stimuli, some of which are presently absent.

Two groups were tested on the binary choice guessing task. One group received methamphetamine and the other soda tablets as placebo; on repeat testing 1 week later the groups were reversed. The subjects were started off on a 50:50 ratio of which two lights would appear but were then changed to a 25:75 ratio. The methamphetamine-treated group changed the ratio of their guesses to the imposed ratio more slowly than the sodium bicarbonate treated group. Their behavior was thought to be the result of the restriction of range of attention to immediate clues and away from those which, being in the past, were peripheral.

Another task which Callaway used to show the attention-narrowing effects of methamphetamine is the Stroop test (1935). This test consists of three cards. On the first of these, names of four colors are printed; on the second are dots of the same four colors. On the third card, names of colors are printed in the inks of the four colors used in the first two

cards, the name of the color being different from the color used to print it. Using Card I the subject has to read the names of the colors, with Card II he has to name the colors of the dots, and with Card III he has to name the color of the ink used to print the names of conflicting colors. It is assumed that performance on Card III will be facilitated by narrowed attention acting to exclude the printed name of the color. The effect of methamphetamine is in fact to improve performance relatively on Card III.

In summary, Callaway states that the procedures which result in narrowed attention, the cold pressor test, and the use of amyl nitrite and methamphetamine, can all be said to have a sympathetic action; on the other hand, he also reports that an anticholinesterase nerve gas has similar effects. The common factor in all the procedures is that they result in an alert EEG. Whereas, however, the amphetamines produce both behavioral arousal (shown in animals by opening of the eyes, movements of the ears, etc.) and EEG arousal (Bradley and Elkes, 1957), anticholinesterases, such as phystostigmine and nerve gas, produce EEG arousal without behavioral arousal. Callaway's suggestion is that narrowing of attention may be the missing behavioral correlate in this case.

As a corollary of these findings it can be suggested that a drug which produces the opposite of EEG arousal should produce "broadness of attention." This possibility was investigated by Callaway and Band (1958) using atropine and Callaway (1959) using amobarbital. Atropine, while producing high-amplitude slow-wave activity in the EEG indicating drowsiness, is reported not to reduce behavioral alertness. It acts as a reticular formation inhibitor but has minimal sedative effects and is not considered to be a cortical inhibitor. Using the Stroop, the Gottschaldt, and the Luchin's water jar tests, results were obtained with atropine which were consonant with the idea of "broadened attention." Amobarbital was found to have a similar effect in the later study although some reservations in the interpretation of the results must be felt because of skewed distributions due to performance near to task ceilings.

The finding of narrowing of attention in a situation likely to produce sympathetic discharge is shown in a study by Kohn (1954). In this experiment subjects read stories in conditions which could be described as threatening, distracting, or neutral. The threatening situation was a stage managed presentation with high-voltage sparks, and the suggestion of an intense shock across the subject's temples, and thus likely to produce an intense emotional reaction. The distracting condition was similar but with a reassurance that no shock would be delivered. The stories were recalled later in a neutral situation. It was found that both the relevant and irrelevant items of the story studied under threat of

shock were less accurately reproduced than those studied under distracting or neutral conditions; however, under threat a greater reduction occurred in irrelevant than relevant items. Thus it was suggested that emotional arousal narrowed the field of attention at the time of the original learning to the relevant points of the story.

Callaway's results with the Stroop test are directly supported by Agnew and Agnew (1963). The state of the subject's drive was manipulated by the use of threat of shock and induction of feelings of failure if above-average performance was not achieved. The effectiveness of this procedure was checked by the finding that heart rate was significantly increased under "high drive" conditions. It was found under these conditions, in contrast to "low drive" conditions, without threat of shock or failure, that performance on the Stroop test was indicative of narrowed attention.

Callaway's studies suggest that narrowing of attention occurs under conditions which increase sympathetic activity or produce an alert EEG. The Agnews' and Kohn's studies lead to very much the same conclusion in that a situation threatening enough to evoke sympathetic discharge produced narrowing of attention.

Bharucha-Reid (1962) suggests that stress of any kind can cut down the intake of stimuli and proposes the term "behavioral anesthetic" for this process. This notion has the same connotation as the "feedback mechanism" of Callaway and suggests that stress produces the means of its own perceptual diminution. While also supporting the notion of narrowing of attention under threat Snygg and Combs (1949) state the rather dissimilar viewpoint that the effect of the threat is to reduce the individual's perceptual field to the area of the perceived threat. It would appear from the experiments described that the effect is best left as an empirical observation whose teleological significance is unclear.

Bharucha-Reid suggests four ways in which the mechanism of the "behavioral anesthetic" might work. First, by an increase in activity in the subcortex the sensory receiving areas may be disorganized selectively, the effect being manifest possibly in a particular area of the cortex representing a single sense modality. Second, even when sensory input is "allowed to be received" it may still not be elaborated for awareness by being received without the presence of adequate secondary reinforcing activity from the subcortex. A third possibility is the suppression of sensory input by the cortex itself, while a fourth is the alteration of cortical activity by the diffuse and specific projection systems acting with a particular time relationship. The trouble with these proposals is the wealth of possible explanations which they offer. We should perhaps make some attempt to delimit more accurately the specific conditions

which bring about the attention-restricting behavior before relying on such a general notion as "stress." While stress may not be the correct term to use for sexual excitement it is of interest to note that Kinsey *et al.* (1953) report not only narrowing of attention in the form of inability to visually perceive actions taking place in the surroundings, but also raising of tactual and pain thresholds during orgasm. Unfortunately, apart from changes in the cardiovascular system, physiological changes accompanying sexual excitement are poorly documented and we do not seem to be entitled to say whether narrowing of attention in these conditions is accompanied by increased sympathetic or fast EEG activity.

Drew (1963), discussing work done during World War II (Davis, 1946a,b; Drew, 1940), reported a gradual restriction of the field of attention during a prolonged fatiguing performance on a task resembling flying. Instruments at the extreme periphery ceased to be responded to, or signals were frequently missed, more so than those nearer the center, until eventually even the central flying panel seemed to be responded to one instrument at a time. Drew reports this behavior as occurring in the context of the subject's increased irritability and presumably therefore increased arousal. In discussing this work, Drew makes an important point. He suggests that the term "tunnel vision" which is the label given to this restriction of the attentive field is a misnomer and an artifact of experimental design. It suggests a restriction from the periphery to foveal areas while attention actually "becomes restricted to that part of the field from which stimuli are expected, whether this be central or peripheral. In the usual experimental design, the main task is presented centrally, with less important or less frequent stimuli to the periphery. Under these conditions, restriction is from the periphery to the centre. If the design is changed, however, the restriction will occur to whatever part of the periphery given the most frequent or most important signals."

The finding of restriction of attention to the central task is shown clearly in an experiment by Bahrick, Fitts, and Rankin (1952). The effect of high and low incentives on performance on a central tracking task with a subsidiary task involving response to intermittent peripheral stimuli was studied. Results showed that the high incentive facilitated performance of the central task but in general interfered with the performance of the peripheral task.

The picture arising from this cursory review of some of the relevant literature is that narrowing of attention occurs under conditions which may be loosely classed as arousing, that is, under conditions that might be considered to produce alert EEG or to evoke sympathetic discharge. This is in essence the position which is put forward in an extensive re-

view by Easterbrook (1959), who says, "It is an empirically derived generalization that, when the direction of behaviour is constant, increase in drive is associated with a reduction in the range of cue use. . . . The term drive refers to a dimension of emotional arousal or covert general excitement which underlies or occurs simultaneously with overt action." These phasic conditions which are productive of narrowed attention in normal subjects should be borne in mind when one is considering perceptual dysfunction in the presence of longer-term disorders of activation, or arousal, which may be said to characterize certain stages of the schizophrenic illness.

A possibility which has been glossed over so far is that we should perhaps consider two forms of narrowed attention. In the case of, for instance, size-constancy experiments and those involving awareness of peripheral visual cues we are concerned with spatial narrowing or narrowing concerned with stimuli present at a moment in time. In the binary guessing game used by Callaway, however, the concern is with temporal narrowing. Whether this latter should more correctly be termed a defect of memory is a matter for discussion but it is clearly connected with such terms as "set" in the Rodnick and Shakow (1940) sense of "maintenance of a set to respond."

B. Experiments on Schizophrenics

1. Spatial

a. Size Constancy. It will be remembered that the postulated link between size-constancy experiments and the phenomenon of alterations in the span of attention is the notion that it is by perception of objects and cues in the peripheral field that the reduced retinal image of an object is corrected and appears to maintain its size when viewed from a distance. If, owing to narrowing of attention, peripheral cues are not perceived, this correction will not take place and the distant object will appear smaller.

A study which directly relates the work on normal subjects, particularly that of Callaway (Section II,A), with studies of size constancy in schizophrenia is that of Weckowicz (1958). In this experiment the schizophrenic patient had to adjust the length of a vertical rod, placed near him, so that it had the same apparent size as that of a distant standard. In addition to this measure of size constancy, the patient's blood pressure response to a mecholyl injection (Funkenstein, Greenblatt and Solomon, 1949) was measured. It was found that "schizophrenic patients in whom there is a rise in blood pressure after a mecholyl injection tend to give a higher estimation of the size of distant objects and therefore

have a higher and better preserved size constancy, while those schizophrenic patients in whom there is little or no rise of blood pressure tend to underestimate the size of distant objects to a greater degree, thus having decreased size constancy." Using the mecholyl test as a measure of sympathetic reactivity (Gellhorn and Miller, 1961), Weckowicz reasons, no doubt on the principle of "the law of initial value" (e.g., Wilder, 1958), that sympathetic reactivity is greatest in those patients with the lowest sympathetic tonus and vice versa. A high level of sympathetic tonus in a group of schizophrenics therefore tends to be related to a decrease in size constancy, and is in accord with findings on normal subjects. Earlier Weckowicz (1957) carried out an experiment on chronic schizophrenics whose clinical picture, after a mean hospital stay of 7 years 9 months, was that of hebephrenia and schizophrenia simplex, although some initial diagnoses were of paranoid schizophrenia. Non-psychotic hospitalized patients and normal controls were also tested. Again the task was that of matching the size of a nearby adjustable rod to that of a standard placed at 7.5 or 15 meters from the subject. In comparison with the other groups schizophrenics underestimated the size of the distant object and did this increasingly so at the greater distance.

This finding of decreased size constancy in chronic schizophrenics is replicated in a study by Hamilton (1963). In this experiment a variety of stimulus objects were used. These objects (standards) were placed 1 meter away from the subject; the variable was 3 meters away and under control of the experimenter. The positions of standard and variable are thus the inverse of those of Weckowicz (1957) but similar to those of Callaway and Thompson (1953). An additional feature was the use of two conditions: one, "cueless," in which judgments were carried out with the stimuli exposed in a darkened tunnel, so that peripheral distance cues were reduced; the other, a "cue condition" carried out in full daylight. Thus if the reduced size constancy of the chronic schizophrenic is due to his inability to perceive peripheral cues because of narrowed attention, his performance should not differ significantly from normal when the normal is deprived of distance cues. The study was carried out on chronic non-paranoid and chronic paranoid schizophrenics, a small group of manic depressives, neurotics, and normal controls. In general it was found that there were no differences in performance between normals and neurotics or among the group of psychotics. There was in general, in conditions where cues were present, a significant tendency for the psychotic subjects to display underconstancy in comparison with normals and neurotics, whose scores tended to approach values for perfect constancy. In the absence of distance cues the size judgments even of normal control subjects were liable to considerable error. In the cueless conditions only two

of the seven objects used as stimuli gave normal-psychotic constancy differences which were significant. It is important to note that these two objects, a penny and a packet of cigarettes of well-known brand, were familiar and of fixed size, while of the remainder of the stimuli, two diamond-shaped cards, a rod, a playing card, and a medicine bottle, only the playing card had a fixed size. Hamilton employed an index of cue responsiveness in which the size constancy in "cue" and "cueless" conditions was compared. It is of considerable interest to note that, for the total psychotic group, reduced size constancy was significantly correlated with low cue responsiveness, as would be expected from the mechanism suggested at the beginning of this section.

In view of the decrease in test intelligence which has been shown to occur as part of the psychological deficit shown by schizophrenics (Hunt and Cofer, 1944) it is of importance to note that Hamilton confirms the finding of Thouless (1932) of a negative relationship between size constancy and intelligence. The results reported do not therefore appear to be a subsidiary effect of the decrease in tested intelligence shown by schizophrenics.

In some contrast to these findings is an experiment by Raush (1952), who investigated the size constancy of paranoid and nonparanoid schizophrenics and a group of normals. While Raush does not give details of the chronicity of his subjects it is evident from such statements as "the diagnosis was consistent throughout the course of hospitalization" and "cases had been reviewed no more than 6 months prior to the experiments" that it was a relatively chronic population that was used. In Raush's experiment the standard stimulus was a disk of light 20 feet away and in front of the subject; the variable stimulus was a disk of light controlled by the subject and placed on a table 4 feet away from the subject and at right angles to the standard. Trials were given in the dark, i.e., cueless conditions, and with room lights full on, i.e., cue conditions. The results of this experiment were that all subjects showed overconstancy in both conditions. However, the only significant differences were those between the paranoid and nonparanoid schizophrenics, and the paranoid schizophrenics and normals in the cueless, dark conditions, the paranoid patients showing higher constancy than either of the other groups. This very different picture from that obtained by the previous investigations is difficult to explain. The general overconstancy of all subjects was thought by Raush to be due to a space error arising because of the 90-degree angle between standard and variable. This angle was 5 degrees in Hamilton's experiment and was of a similar order in the experiments by Weckowicz and Callaway. Holway and Boring (1941), using a similar orientation of standard and variable

to those of Raush's, obtained a larger degree of overconstancy with their normals than did Raush. This finding is in many ways similar to that of Klein (1954), who found that size judgment from memory gave consistent overestimations in comparison with conditions where size judgments were made with the standard in the peripheral field. Clearly the process of comparing stimuli at right angles as in Rausch's experiment involved memory, however immediate. Another puzzling feature is the fact that differences which are shown are found in the cueless condition; however, while the normals and nonparanoids give slightly less constancy in the dark than the light conditions the overconstancy of the paranoids is increased under these conditions. The main contrast between this and the Hamilton experiments, given that the experimental conditions brought about a general bias to overconstancy, is the finding that the paranoid schizophrenics behave differently from the nonparanoids. This finding of exceptional behavior by paranoids will be discussed in Section V.

It is perhaps noteworthy that Raush's schizophrenic population was on the whole younger (about 25 years old) than Hamilton's (40 years) or Weckowicz's (36 years). Not only therefore are they likely to be less chronic, but according to Raush their diagnosis had not changed. Weckowicz noted that although some patients in his experiment had been diagnosed as paranoid on admission this diagnosis no longer applied. The point that is being made here is not that age in itself affects size constancy—indeed Hamilton has shown no differences with age—but that age may reflect differences in chronicity and possible changes in diagnosis. Hamilton showed no changes in size constancy with chronicity when his group of patients was divided at more or less their 5 years stay in hospital. It can be suggested, however, that the change in physiological state which parallels an acute-to-chronic change occurs very early in hospitalization, and by division of his patient group at more or less than five years stay Hamilton would be unlikely to show any differences.

A criticism of Raush's study which is made by Lovinger (1956) is that Raush used patients who were in good contact with reality so as to obtain optimal conditions for size-constancy testing. This selection would tend to bias the results in the direction of not finding reduced size constancy. Lovinger therefore carried out a study in which two groups of schizophrenics in good and in poor contact were selected. The selection of patients was made on the basis of ratings of contact and whether they lived in open or closed wards. Using illuminated disks as targets in line and not a right-angular arrangement of standard and variable, Lovinger found that "under experimental conditions involving minimal distance cues schizophrenics in poor contact manifested less size constancy than either schizophrenics in good contact or normals." It is of

importance to note that Lovinger used three cue conditions: (1) maximum, with full lighting, (2) minimal, with reduced lighting, and (3) no cues, darkness. In both the last and the first conditions no differences due to diagnosis were found, probably because even the poor-contact schizophrenics noticed peripheral cues under full lighting, and on the other hand, in darkness the normals were as handicapped as the schizophrenics. The possible intrusion of distance cues even for out-of-contact schizophrenics would seem to be a likely explanation for the finding (Leibowitz and Pishkin, 1961) of fully maintained size constancy in a group of chronic schizophrenics chosen as very withdrawn. The experiment was carried out under fully lit conditions, which might have resulted in extreme prominence of distance cues.

Probably related in a loose fashion to the degree of reality contact used as a factor in Lovinger's experiment is thought disorder. Weckowicz and Blewett (1959) started from the notion that paranoid schizophrenics who show little thought disorder also show a well-maintained size constancy (Raush, 1952), while schizophrenics who show poor size constancy are also likely to be thought disordered. Using 40 chronic schizophrenics as subjects they found that there was a positive correlation between poor size constancy and imparied ability for abstract thinking. Weckowicz and Blewett interpret their result in terms of an inability to "attend selectively" or "to select relevant information" which is responsible for both size constancy and thought disorder. This interpretation of size constancy would seem to be at variance with the ideas of "decreased general awareness of the gradients by which the shrunken retinal image is corrected" which is said (Weckowicz, 1957) to result in decreased size constancy. The relationship of narrowed attention to thought disorder will be discussed later in the context of overinclusion.

A study which has been included in this section because of its very close relationship to size constancy is that of Hartman (1962). This work is on the apparent size of after-images and is a rough converse of size-con-stancy experiments in that there is a fixed retinal image size which accord-ing to Emmerts law changes its phenomenal size with perceived distance. Quoting experiments by Edwards (1953) and Boring (1940), Hartman suggests that if all cues to distance are removed from the viewing situa-tion the size of after-image will depend on its retinal size and will thus remain constant. Two groups of schizophrenics and two normal groups were tested in this experiment; one group of patients was categorized as delusional and the other as nondelusional. In both cases the length of hospital stay was about 4 years. While the scanty reporting of actual details of results makes it difficult to discover actually what happened it was reported that "chronic delusional schizophrenics differed from

chronic nondelusional schizophrenics and normals in their perception of size changes of the after-image, i.e., they were less likely to respond as would be predicted on the basis of Emmerts law." No subject, normal or schizophrenic, was able to report after-image changes as distance varied when only monocular cues were available; convergence cues, however, did bring about a difference in after-image size but less in delusional patients than other subjects. These cues were only operative at short distances. While showing that paranoid patients are apparently different from the rest, the light which this single study throws on the size-constancy findings is unfortunately small, apparently because few distance cues were available for even normal subjects to use.

b. *Distance Constancy.* Callaway and Thompson (1953) failed to produce evidence that distance judgment was affected under sympathetic discharge evoked by the cold pressor test, and thus dismissed this as a factor responsible for changes in size constancy under stimulating conditions. Weckowicz, Sommer, and Hall (1958) carried out an out-of-doors distance judgment experiment which showed that chronic schizophrenics underestimate distance. Schizophrenics, when asked to judge successive equidistant intervals, had a steeper perceived gradient and reached zero sooner than normals. The perceptual field of the chronic schizophrenic is apparently lacking in depth, he lives in a "flatter world." Size-constancy scores obtained in a similar way to Weckowicz's other experiments were available for the patients in his experiment. It was found that poor size constancy was significantly correlated with poor distance judgment. This correlation has been criticized by Rump (1961a,b) on the basis of the lack of validity of the "invariance hypothesis" of a reciprocity between perceived size and perceived distance. Hamilton (1963), who asked his subjects to make distance judgments at the end of his size-constancy experiments, agreed with Rump's criticism of Weckowicz et al., insofar as he found almost wholly negative and nonsignificant correlations between distance judgment and size-constancy measurements carried out on the *same* apparatus in contrast to the separate determinations of Weckowicz et al. Hamilton's results do, however, show, in agreement with Weckowicz et al., that nonparanoid schizophrenics underestimate distance, but also that paranoid schizophrenics markedly overestimate distance. It was also found that cue conditions benefited the distance judgments of nonpsychotic subjects more than those of the nonparanoid schizophrenics. It might thus appear that distance judgment deficiencies are possibly interpretable in terms of narrowed attention, and hence while full appreciation of peripheral cues is a necessary condition for good size and distance judgments the mechanisms responsible for each are independent. An experiment carried out by Nelson and Caldwell

(1962) gave no indication of an inaccuracy of depth perception in acute schizophrenics. This experiment was, however, carried out in a rather different way from those earlier and involved the positioning of a variable target to have the same apparent distance as a fixed target, an apparently easier task than, for instance, making a judgment of distance in feet (Hamilton) or the doubling or halving of standard distances (Weckowicz *et al.*). Although it was not significant, there was a tendency for the acute schizophrenics in Nelson and Caldwell's experiment to be more accurate than normals when using a post as a stimulus, possibly indicating some degree of overconstancy in these acute patients. Other objects which might have acted as affective stimuli were used, pictures of a woman, dog, man, or circle, but while there was some tendency for normals to show an order of placement of items this was not shown by schizophrenics.

  c. Perception of Relevant and Irrelevant Material. Venables (1963a) carried out an experiment concerned with the degree to which schizophrenics perceived irrelevant stimuli with the direct intention of testing the narrowed-attention hypothesis. A card-sorting task was used in which cards had to be sorted on the basis of the presence of one or another of two relevant letters in the context of eight irrelevant letters. Four sortings were made by each subject and then without being informed of any change he was asked to make a fifth sorting on the basis of the same relevant letters but in the context of a different set of irrelevant letters. On the basis of each individual regression line summarizing the trend for the second, third, and fourth sorting, a time for a fifth sorting, had the irrelevant letters not changed, was estimated. This was compared with the observed time for the fifth sorting. It was suggested that the degree to which the subject had paid attention to the irrelevant letters would be reflected in the disturbance shown by the difference between the estimated and observed times for the fifth sorting. Following Callaway (Section II,A) it was thought that narrowed attention would be related to the degree of cortical activation shown by the subject. As an index of cortical activation the subject's threshold of fusion of paired light flashes was measured. Strictly speaking the two-flash threshold should be thought of as a measure not of cortical activity but of cortical responsivity or of temporal resolution. Good temporal resolution, which involves short cortical recovery time, is related to the refractoriness of the cortex. It is evident that a second stimulus may not give rise to a cortical response if it falls into a period of complete refractoriness after an initial stimulus. The shorter this period of refractoriness the better the temporal resolution. The ability of the cortex to resolve pairs of stimuli which are close

in time appears from the experiments of Lindsley (1958), Steriade and Demetrescu (1962), and Schwartz and Shagass (1963) to be under the control of the reticular formation whose ascending activating function is usually thought of as increasing cortical activation. Thirty-four chronic schizophrenic patients were tested and their two-flash fusion thresholds correlated with the degree of card-sorting disturbance used as a measure of narrowed attention. It was found that high cortical responsivity was significantly correlated with narrowed attention, thus supporting Callaway's findings. It had been found previously (Venables and Wing, 1962) that there was a relationship between two-flash threshold and a rating scale measure of withdrawal (Venables, 1957) such that high cortical activation was found in the most withdrawn patients, if these were non-paranoid or incoherent paranoid schizophrenics. Narrowed attention could thus be expected to be related to the degree of withdrawal shown by the subject. For the 26 nonparanoid patients in the sample this relation was significant. However, no relationship was evident for the remaining eight coherently paranoid patients. Similar data relating withdrawal and narrowness of attention showed a positive relation between the two measures in another group of nonparanoid patients and no relationship in a larger paranoid group. These results should be compared to those previously reported where for instance Lovinger (1956) found reduced size constancy in a group of patients having reduced contact with their environment and hence most probably comparable to the withdrawn patients in Venables' study. The lack of relation between withdrawal and narrowness of attention in paranoid patients also is in line with the findings previously reported of differences between these patients and other schizophrenics. The need to distinguish those paranoid patients having coherently expressed delusions from those whose delusions are expressed incoherently and who behave in many ways like non-paranoids (Venables and Wing, 1962) may account for the lack of difference between paranoid and nonparanoid schizophrenics in Hamilton's (1963) study.

The idea of stimuli as relevant or irrelevant to the task in hand is at the basis of studies of incidental learning. Greenberg (1953) had a group of 44 chronic schizophrenics and 44 nonpsychiatric patients perform three experimental tasks: (1) "Color-Position" in which the recall of the colors of geometric forms was the directed or intended task and the recall of the position of the form the incidental or undirected task, (2) "Paragraphs," in which the recall of the content of one paragraph from the Wechsler Memory Scale was the intentional task and the recall of the second paragraph the incidental task, when the instructions called

only for tallying the frequency of words, (3) "Metal-Nonmetal," in which the recall of familiar metal items was the directed task and the recall of familiar nonmetal items, exposed simultaneousy with the metal items, was the undirected task.

The results showed that the nonpsychiatric group was better than the schizophrenic group on all measures of incidental learning. Even after correction for differences in intentional learning by analysis of covariance, the normal group was superior to the schizophrenic on two of the three incidental learning scores. It is important to note that Green-berg suggests that the conclusion that schizophrenic patients show a greater impairment in incidental than in intentional learning "must be limited to the category of patients tested in this study since there are some indications that it does not apply to other categories such as para-noid and acute schizophrenia."

Further support for the contention that incidental learning in chronic schizophrenics is less than that in normals is given by a study of Topping and O'Connor (1960) in which there is a tendency for non-paranoid subjects to give lower incidental learning scores than normals, with a paranoid group occupying an intermediate position. A study by Winer (1954) showed lower than normal incidental learning in hebe-phrenic and catatonic schizophrenics, while paranoid schizophrenics had scores indistinguishable from normal. A further study which can be loosely included under the present heading is that of Dunn (1954). The experiment was carried out on 40 chronic schizophrenics and 40 non-psychiatric hospitalized controls. Four types of scene were exposed briefly to the subject, including three "mother and son" pictures showing "whipping," "scolding," and "feeding." A neutral picture having the same relative sizes of figures showed a house and a tree. There were five versions of each picture, each having differing steps of detail change, such as the position of a branch in the tree-and-house pictures. The subject's task was to report whether there was any difference in a pair of pictures or whether they were the same. It was found that in the whipping and scolding pictures the schizophrenics were significantly worse than normals in the observation of detail; this difference did not reach significance for the feeding and house-and-tree pictures. While this is not Dunn's interpre-tation these results may be viewed as supporting a view of a narrowing of attention in chronic schizophrenics exaggerated by the emotion evoked by the whipping and scolding pictures. It may be, however, that an emotional reaction is evoked by these pictures in all subjects and it is only the combined effect of this on top of an already aroused schizo-phrenic "habitus" that is sufficient to produce the narrowing of attention which brought about nonperception of detail in the pictures.

2. Temporal

a. Perception of Statistical Structure. So far we have discussed experiments interpretable in terms of narrowing of attention that is spatially orientated. In this section we turn to experiments which can be interpreted in terms of temporal narrowing. One of the tasks which Callaway (Section II,A) used to test the hypothesis of narrowing of attention under autonomic stimulation was that of binary choice guessing. He showed that subjects receiving methamphetamine changed the statistical structure of their guesses to that of the reinforced answer pattern more slowly than a group not receiving methamphetamine. The proposed mechanism of this phenomenon is that past answers and guesses, being in the periphery of attention, are not so readily perceived under conditions of narrowed attention and are not responded to. Changes in guess pattern do not then take place so readily.

A somewhat similar experiment was carried out by Ashman (1959) on a group of schizophrenics who were described as first admissions with a median age of 35 years. The length of stay in this first admission was not described, but unless the group contained a large proportion of paranoid patients it might be suggested that at 35 years the group be considered comparatively chronic.

The total group was divided into five subgroups, each having a different pattern of initial, interpolated, and final tasks. The initial and final tasks all consisted of 100% reinforcement for one response, while the interpolated tasks had different patterns of 50:50 reinforcement. Ashman summarized his results as showing that "schizophrenics do not appear to be able to utilize prior probability learning on responding appropriately to later patterns." They tend to perseverate to a reinforced response after the cessation of a reinforcement renders the response inappropriate in terms of a different pattern of events.

Employing a more complicated statistical learning task the present author carried out some preliminary studies on this problem which were eventually abandoned because the task proved too difficult. However, results as far as they went throw light on the problem. The task in this instance was a four-choice sequential probability learning task with probabilities of 0, 17, 33, and 50% of the appearance of any of the four lights following the appearance of any one. The probability pattern was complete on 64 trials and four blocks of 64 trials were used with each subject. Six nonparanoid chronic schizophrenics were tested and as a measure of their activation level, skin potential level was measured (Venables, 1963b). It was found that the three patients with low skin potential levels learned the task reasonably well, whereas the three

patients with high skin potential levels, and hence possibly narrowed atten-
tion, could not learn the pattern and resorted to position habits. It
could be argued that as the three patients with high skin potential levels
were also more withdrawn, and hence possibly more sick than the other
three patients, their poor performance resulted simply from the fact
they were less able to undertake the task. This did not seem to be the
case, their performance gave the impression more of consisting of short
runs of trials of fresh patterns which, not being reinforced, eventually
led to position habits. On the other hand the three patients who learned
achieved performances progressively closer to the imposed pattern.

Closely associated with this type of experiment are those on language
when the main concern is with its statistical structure. Lewinsohn and
Elwood (1961) studied the learning of word lists having different orders
of approximation to unconstrained language as in the Miller and Selfridge
(1950) experiment. Four groups of subjects were tested, acute and chronic
schizophrenics, nonschizophrenic patients, and normals. It was found that
the chronic schizophrenic patients stood out from the rest in their inability
to make use of the contextual constraints of language. When the vocabu-
lary level was controlled it was found that the difference between chronic
patients and the rest disappeared; however, it could be said that vocab-
ulary performance insofar as it is the selection of a verbal description of
a word while keeping that word in the forefront of attention might very
well be contaminated by the same mechanism of narrowed attention.
The confounding effect of vocabulary level is found in experiments by
Pearl (1963), who examined the performance of process and reactive
schizophrenics on a version of Shannon's (1951) guessing game. Shannon's
method assumes that people have an implicit knowledge of language
statistics and when required to guess the next letter of a meaningful lan-
guage sequence, use this information to predict the most likely letter.
Pearl found that if the total population of acute schizophrenics which
he tested was dichotomized on the basis of vocabulary level, there was
no difference in performance of the high-vocabulary process and re-
active patients. Low-vocabulary process schizophrenics, however, took
significantly more guesses to arrive at the correct content of a sentence
than low-vocabulary reactive patients, and both these groups were in-
ferior to normal subjects with similarly low vocabularies.

These language experiments tend to support the general notion that
the selection of the immediate response which is in the foreground of
attention is not helped in chronic or in process schizophrenics by the
pattern or structure of the words or letters normally surrounding it.

b. *Effects of Anchor Stimuli.* Weinstein, Goldstone, and Boardman
(1958) carried out an experiment on the time estimation of schizophrenics

in which they confirmed the finding of Lhamon and Goldstone (1956) that schizophrenics tend to overestimate the duration of a clock second; their main concern, however, was to determine the effect of anchor stimuli upon performance. They found that both normals and their sample of chronic schizophrenics responded to the "pulling" effect of long- and short-interval anchors. That is, if the series of time intervals used in the determination of "Second Estimate Point" (SEP) started with a short interval the SEP was "pulled" down, whereas if the series started with a long interval the SEP was "pulled" up. However, if a subsequent series was given with the anchors reversed it was found that the judgment of normal control subjects was influenced by the earlier series whereas the schizophrenics tended only to respond to the effect of the immediate series. It appears therefore that the range of attention of these chronic patients was narrowed and did not involve the use of cues presented earlier in time.

Salzinger (1957) studied the effect of a prior experience of lifting a heavy or a light weight on the judgment of the categories of a series of five weights. In this experiment the performance of acute rather than chronic schizophrenics was compared with that of controls. It was found that after the prior experience of the heavy weight the schizophrenics shifted *more* than the controls. Performance did not differ after the light anchor. Thus it would appear that consonant with the idea of increased awareness of objects in the environment or broadness of attention in the acute schizophrenic, the patients were more aware during their subsequent judgments of the impression left by the earlier perceived heavy weight and then judgment was affected.

An experiment carried out by Boardman, Goldstone, Reiner, and Fathauer (1962) on the effect of anchor stimuli on the judgment of a length of one inch provides further evidence to support the earlier studies. Groups of acute and chronic schizophrenics and normals were each divided into three subgroups. The first made size judgments using a modified method of limits technique starting with a neutral 1.0-inch anchor; the second started with a 2.0-inch anchor and after a brief rest followed with a series with a 0.1-inch anchor. The procedure for group 3 was the reverse of group 2. The comparisons of greatest interest are those between the judgments for series with 0.1-inch anchors presented first or second, and with 2.0-inch anchors presented first or second. It was found that neither the normal nor the chronic schizophrenic group showed any significant differences between their initial or final 0.1-inch or 2.0-inch anchor judgments, whereas the effect of the previous anchor situation was marked in the judgment of the acute schizophrenics. The effect of the immediate anchor condition was felt when it was 0.1-inch in

all groups but the immediate effect of the 2.0-inch anchor was felt only in the acute group. It is tempting to suggest that owing to "tunnel vision" the extremities of the 2.0-inch anchor were not perceived by the chronic schizophrenics but this explanation seems too facile. We are thus presented with further evidence that the acute schizophrenic still has in his broadened attention the impression left by the previous anchor given while this is excluded from the attention of the chronic schizophrenic. The lack of effect of anchors on normal subjects in this study confirms the results of an earlier study (Boardman, Aldrick, Reiner, and Goldstone, 1959) and suggests that "the concept of an inch is a social standard of magnitude and is stable in a normal population."

3. Discussion and Conclusions

It is a pitfall of any review which attempts to gather together a rather disparate set of findings under some blanket hypothesis that some violence is done to the original findings. It is hoped that this has been avoided in the present section and moreover that there has not been unwitting selection of data to fit the thesis.

In summary, what appears to have emerged from the experiments discussed is the finding that on the whole chronic schizophrenic patients behave in a way which suggests that their attention is narrowed. Within this group the extent of the narrowing is related to the patient's withdrawal, or the degree to which he appears to be in contact with his surroundings. There is some evidence that paranoid patients may stand as an isolated group who behave differently from other chronic patients. In addition there is much slighter evidence that patients called either "poor premorbid" or "process" schizophrenics tend to act similarly to chronic patients, and that "reactive" or "good premorbid" patients tend to act with patients classed as acute as though they conformed to the clinical picture of overawareness of their surroundings. This suggestion, however, is much more tenuous than the narrowing of attention in chronic subjects. Following from these results is the collateral proposal resulting from Callaway's findings that narrowed attention is manifest in conditions which produce cortical alertness; that cortical overactivity is to be expected in chronic schizophrenics and the degree of overactivity is related to the patient's mental state.

Another point which has emerged is that the effect of narrowed attention can be felt both in respect to the sensory stimuli incoming at any one time and on items which are not contemporaneous. Thus it is suggested that items which are the more remote members of a set, some of which are presently attended to, may be eliminated from attention by narrowing, thus affecting the "meaning" of the items in the

focus of attention. The use of the word "meaning" has evident relevance when one is considering, for instance, the statistical structure of language, but it is just as relevant when we take the view that the meaningfulness of the organism's surroundings is in relation to the degree to which they are perceived as part of a geographic and historic whole.

Close parallels may be drawn between the position reached here and that, for instance, put forward by Osgood (1957). Following the postulates of Hull (1943) Osgood suggests that the effect of drive is to reduce the number of alternative responses in a hierarchy. This comes about because of the multiplicative relationship according to which drive increases the probability of alternative reactions in proportion to their initial habit strengths. Thus, under high drive dominant responses become more dominant and less probable responses in the hierarchy become even less probable. If this is applied (as Osgood does) to language behavior, the number of associates arising from the presentation of a particular word decreases under high drive, and thus meaningfulness in the sense used earlier decreases. This neo-Hullian picture puts narrowing under high drive at the response end of behavior but the parallels between it and perceptual narrowing deserve further attention. In the case of temporal narrowing the division between the two pictures is largely verbal and revolves around the use of terms such as associations, associative responses, and items in the attentive field which are the result of past experience rather than due to present sensory input. It should be noted that this learning theory viewpoint is in contrast to that which is used, for instance, by Mednick (1958) where high drive is said to act impartially upon correct and incorrect responses and therefore tend to push many irrelevant responses above threshold.

This theoretical confusion is possibly resolved by a suggestion of Callaway and Stone (1960) to the effect that "arousal will lead to (1) reduced probabilistic coding (i.e., increased over-all uncertainty with regard to the stimulus ensemble under consideration); (2) increased filtering (as in our previous concept of narrowed attention); and (3) reduced size of stimulus ensemble under consideration." As Callaway and Stone point out, this leads to some ambiguity as it is impossible to tell a priori whether a peripheral stimulus will be excluded because of narrowness of attention or included with nearly the probability of the most probable stimulus because of prediction (1). This model does comply with the Hullian prediction used by Mednick in which competing responses are elevated nearer to threshold, but is not very useful as an explanatory concept until the conditions for narrowing of attention and the reduction to equality of the remaining probabilities are examined by experiment.

III. Relationship to Other Groups of Experimental Studies

A. OVERINCLUSION

So far, apart from a brief reference in Section I,A, no mention has been made to the extensive work on overinclusion which has obvious relevance to the present review. This has been done deliberately, partly because it would be unnecessary duplication to review the literature on this topic when this has already been done by Payne, Matussek, and George (1959) and Payne (1961), and second because to some extent the findings on overinclusion provide an independent check on the position which has been reached as a result of the present review of an entirely different set of experiments.

Although they do not proceed to test it by other than the more conventional tests of overinclusion, Payne *et al.* (1959) make the suggestion already quoted that overinclusion is an attention defect and is due to the breakdown of a filter mechanism that "cuts out or inhibits the stimuli, both internal and external, that are irrelevant to the task in hand" or, in our terminology, that it results from excessive broadness of attention.

Payne and Friedlander (1962) and Payne (1962) summarize the findings on overinclusion as follows:

Overinclusive thinking may be a relatively specific disorder, independent of general intelligence; it influences performance on a number of different types of tests, yielding a clear factor when intercorrelations between tests are factor analyzed. It is confined to patients diagnosed as schizophrenic and among these is not found in retarded patients. Chronic schizophrenics do not appear to be overinclusive, but the abnormality does appear in patients with delusions. The findings suggest the idea that overinclusive thinking may be a relatively good long-term prognostic indication.

Using the object classification test (Goldstein and Scheerer, 1941), Payne (1962) presented data from five groups of subjects. On this test acute schizophrenics stand out from the other groups of normals, neurotics, depressives, and chronic schizophrenics in showing a greater degree of overinclusive thinking. The chronic schizophrenic group is almost completely within the normal range of scores for numbers of unusual sortings.

The same picture of overinclusive behavior occurring only in acute schizophrenics is found when the scores from three tests are combined (Payne and Friedlander, 1962).

Feinberg and Garman (1961) studying thought disorder in schizophrenics with a modified version of the Progressive Matrices test

categorized subjects' responses in terms of their plausibility. Acute schizophrenics made many errors categorized as "implausible" in that their responses showed no geometrical similarity to the stimulus patterns. Chronic schizophrenics, on the other hand, made few errors of this kind and it was suggested that their performance should be interpreted in terms of "underinclusion" rather than "overinclusion."

These studies tend to support the conclusions drawn from the examination of data from experiments not concerned with testing over-inclusion (Section II,B) that the perception of the acute schizophrenic is such that there is little selection of relevant and irrelevant material; it is in other words broad, while that of the chronic schizophrenic is narrow. That overinclusion occurs in schizophrenics with good prognosis or those classed as reactive also fits the ideas put forward earlier in this section that overawareness of surroundings is to be found in "good premorbid" patients.

B. STIMULUS GENERALIZATION

Mednick's (1958) learning-theory approach to schizophrenia has as one of its main tenets the notion that acute schizophrenics show an abnormal degree of stimulus generalization. Defined as the condition "when a response, having been trained to a stimulus, is also elicited by similar stimuli," stimulus overgeneralization is a close conceptual counterpart to overinclusion. If we consider that excessive width of attention is a disorder in which many items have equal prominence in the organism's consciousness and may therefore have equal likelihood of eliciting a response, the closeness of this concept to stimulus over-generalization is apparent.

The work on generalization in schizophrenics does not in fact enable very strong conclusions to be drawn. Mednick (1955) used schizophrenics whose length of hospitalization was over 9 years on a task in which response to a central stimulus light was spatially generalizable to more peripheral lights. He found support for a hypothesis that schizophrenics would show greater generalization than normal. It was not, however, conclusive. On the basis of comparison with work on overinclusion and breadth of attention it should not be expected that the chronic schizo-phrenics used by Mednick would show generalization; however, the fact that half his sample were diagnosed as paranoid would possibly give rise to the small over-all amount of generalization that was shown. Garmezy (1951), however, using a sample of acute schizophrenics, showed that these patients tended to generalize from a training tone to other stimuli along the pitch dimension.

Knopf and Fager (1959) showed some degree of spatial generaliza-

tion in schizophrenics, but as the chronicity of the patients was not designated, the value of their results is minimized. Other findings on generalization in schizophrenics tend to be by inference from experiments on verbal learning. Mednick and Devito (1958), for instance, showed that, in a complex learning situation with competing intralist associations where overgeneralization might be expected to worsen performance, acute schizophrenics were more impaired than chronic patients. Carson (1958), also working from the expectation of greater generalization in schizophrenics and hence greater difficulty with rote learning involving increasing intralist similarity, found his schizophrenic group produced results opposite to those predicted. In the light of the present thesis it is interesting to note that Carson's patients were of chronic status.

On the basis of these rather limited results therefore it seems reasonable to conclude that acute schizophrenic patients may show an excess of stimulus generalization, while this is not the case with chronic patients. As these results seem to be parallel and similar to those on broadened attention, it does not seem wise to draw the conclusion that high stimulus generalization in acute patients arises because of high drive in these patients; rather, the reverse seems likely. Generalization on the basis of high drive as suggested by neo-Hullian theories may profitably be distinguished from generalization arising from broadness of attention brought about by decreased arousal. The whole of this conceptual tangle deserves further experimentation.

C. Sensory Deprivation

The impression gained from experiments reviewed earlier is that acute schizophrenic patients are under constant bombardment from an excess of unrestricted sensory stimulation, whereas the reverse position is the case with chronic patients whose narrowed attention restricts the amount of sensory input consciously perceived. These are changes brought about by the physiological state of the patient; the amount of sensory material available to consciousness can, however, be manipulated externally and by restriction might be expected to produce at least some of the symptoms of chronic schizophrenia in normal subjects. As a corollary, restriction of sensory material might be expected to improve the state of the acute patient.

In a review of anecdotal and experimental material on sensory deprivation in normal subjects, Rosenzweig (1960) says, "One cannot help noting throughout these reports. . . . that there appear again and again Bleuler's cardinal symptoms of schizophrenia; disturbance of associations, disharmony of affect, autism, ambivalence. We see disruption of secondary thought processes, regression to the primary process,

impairment of reality testing and in addition such accessory symptoms as distortion of body image, depersonalization, delusions, hallucinations." That these pathological effects arise not merely from diminution of sensory input per se but rather from the diminution in the meaningfulness of sensory input is shown in a study by Davis, McCourt, and Solomon (1960), who restricted the usual patterning of sensation but supplied input in the form of a randomly flashing light. Discussing the pathological disturbances arising from this procedure they state, "It appears that what the brain needs for normal functioning is not quantity or change of sensation per se, but a meaningful contact with the outside world."

Unfortunately experiments to test the other side of the coin are inconclusive. Gibby, Adams, and Carrera (1960) reported an enduring change for the better in a mixed group of patients containing some schizophrenics of unspecified chronicity after exposure to a sensory deprivation situation. Improvement was also reported by Harris (1959), particularly for hebephrenics, but the change did not last beyond the experimental situation. No change was reported by Smith, Thakurdas, and Lawes (1961), who used chronic patients, or by Cleveland, Reitman, and Bentinck (1963). The lack of change in chronic patients might of course be expected as they might be considered to be already in a state of sensory deprivation. More conclusive evidence from this source must await experiments with better attention to diagnostic and chronicity criteria.

D. "SET" IN REACTION TIME

One of the largest sections of integrated experimental work on schizophrenia is that on the effect of time uncertainty in RT. This has been discussed by Shakow (1962, 1963). A related set of experiments are those on stimulus uncertainty which have recently been reviewed by Sutton and Zubin (1963). While this work deals with "expectancy" or "set" and is thus not strictly on the topic of the present review, some of the statements which have been made in the course of theoretical discussions make it evident that it should be considered at the present time. As an example, we find in Sutton and Zubin (1963) that "studies strongly suggest that the state of readiness of the patient is disproportionately affected by events which are recent in time"; however, on the other hand, these authors say "it is also clear in the foreperiod experiments, and suggested in recent work on modality experiments that long-range influences operating over the whole series of trials act to 'over influence' and 'over prepare' the schizophrenic subject." There thus seems superficially to be a conflict between these studies, which have mainly been

carried out on chronic patients, and those already reviewed, which show temporal narrowing in these patients. In attempting to resolve the conflict it is important to keep in mind the different processes involved in perception and in the development of "sets" or "expectancies." In perception what is always in the forefront is the present stimulus and the narrowness or broadness of attention must be considered in relation to this present anchor point. In the case of expectancy we have a state of the organism built on the basis of past stimuli and experiences and having relevance to response to a future stimulus. The range of relevant items in the attentive field is not anchored to a present stimulus as in the perceptual process. Thus in expectancy we may have the position of the range of attention being narrowed to a particular subset of items in the past so that the subject acts with response to this "minor set" in a manner which is inadequate. For efficient performance attention to a wider range of items is required to develop a "major set" which may bring the range of expectancy up to date. By such an explanation as this the theoretical positions arising from "perceptual" and "expectancy" experiments may be mutually reconciled.

The leading finding of the RT experiments on this subject is that on the effect of regular and irregular foreperiods on speed of reaction. When, in the regular procedure, the foreperiod is kept constant for a block of trials, normal subjects appear to be able to take advantage of this procedure for foreperiods up to 20 sec. Chronic schizophrenics, on the other hand, are only able to benefit from the regularity for foreperiods up to 5 or 6 sec., after which they perform less well than when foreperiods are presented randomly, giving the subject no opportunity to develop a set to respond. This difference in the degree to which subjects were able to utilize the regularity of foreperiod was embodied in an arbitrarily constructed set index (Rodnick and Shakow, 1940), which was found (Rosenthal, Lawlor, Zahn, and Shakow, 1960) to be closely related to a rating of the patient's mental health.

Of particular interest to the present review is a study by Huston and Singer (1945), who examined the effect of sodium amytal on a group of patients whose duration of illness was 9 months and might therefore be considered to be in the early chronic stage. Without drugs these patients showed the same pattern of relationships as those shown by Rodnick and Shakow (1940). The normal subjects had faster reactions with regular foreperiods than with irregular foreperiods up to 20 sec. in length. The schizophrenics, on the other hand, showed that at about 6-sec. foreperiod length the advantage of the regular foreperiod was lost. Under sodium amytal, which it will be remembered (Section II,A) is supposed to induce a broadening of attention and therefore in the

Huston and Singer experiment should enable the subject to benefit from constant foreperiod length, the pattern of schizophrenic results was similar to that of the normals. In a later experiment Huston and Senf (1952) studied the effect of amytal upon the performance on a similar RT task with both early and chronic schizophrenics. Because only three preparatory foreperiods were used results were not strictly comparable to those above. The trend, however, was with the regular procedure under drugs for the chronic schizophrenics to show distinct improvement in comparison with the undrugged condition. The improvement was not marked in the case of the early schizophrenics.

Another study from Shakow's group (Zahn, Rosenthal, and Shakow, 1963) very nicely demonstrates the point under discussion and has particular relevance to the experiments on anchor stimuli (Section II,B,2,b). Using only irregular foreperiods, these experimenters were able to show that when the foreperiod before the one presently involved was longer than the present one the chronic schizophrenic patients were more affected by it than normals. The conclusion is thus reached that, "instead of basing their pattern of preparation on their experience with the series of preparation intervals as a whole, the patients seem to base it, much more than do the normal subjects, on the most recent event in the series."

In the case of stimulus uncertainty experiments (Sutton and Zubin, 1963) the uncertainty is not when the stimulus is going to appear but what stimulus is to appear. In these studies disturbance of performance is shown more by changes of stimulus from light to sound than by the degree to which expectancy for one or another type of stimulus is built up. It is with this sort of influence of the parameters of the stimuli within the total context that the next section will be concerned.

IV. Stimulus Parameters as Determinants of Input Dysfunction

A further set of factors which are responsible for dysfunctions of input in the schizophrenic patient are those concerned with the simple parameters of stimuli. Other more complex parameters such as their aversive or social content when stimuli act as reinforcing agents have been reviewed recently by Silverman (1963) and will not be considered here. Data are available in any quantity only on the effects of the parameters of modality and intensity and the discussion will be limited to these aspects. It might be expected, however, that the proposals put forward by Kaplan (1960) on the differential sensitization to long and short wavelengths by autonomic arousal would provide a further dimension on which normal-schizophrenic differences might be expected.

A. MODALITY

In Section II,A some proposals by Bharucha-Reid (1962) concerning the way in which subcortical dysfunction may lead to disorganization at the cortex were outlined. One of these was that the subcortex may affect the function of sensory receiving areas selectively and thus possibly affect a single sense modality and thus disturb the normal relationship which exists between the different modalities.

There is some evidence for such a disturbance in schizophrenia. Venables and O'Connor (1959) carried out an experiment on the reaction time to auditory and visual stimulation, using groups of chronic schizophrenic subjects divided on the basis of rating scales measuring withdrawal and paranoid tendency. A control group of normal subjects was also tested. The finding usually reported (cf. Teichner, 1954) that auditory RT's are faster than visual RT's with normal subjects was replicated in the case of the normal subjects in this experiment. It was also shown by those schizophrenics designated as both paranoid and not withdrawn. (This is a group similar to that group labeled coherent paranoids which behaved differently from the remainder of the patients in a study detailed earlier—Section II,B,1,c.) The other schizophrenics had faster reactions to visual stimuli than to auditory stimuli. This result is in conformity with a much earlier study (Wells and Kelley, 1922) in which a group of dementia praecox patients showed a tendency for RT to sound to be relatively longer than to light as compared with normals. The effect of different stimulus modalities was also shown in a study of the effect of background stimulation on the skin potential response of chronic schizophrenics (Venables, 1960). It was found that an auditory background had a greater effect than a visual background in slowing down responses in a nonwithdrawn paranoid subgroup, while in the remainder of schizophrenics the tendency was for the response to be speeded up. The ability of the cortex to resolve pairs of stimuli presented in close temporal proximity has been suggested as a measure of its excitability (Venables, 1963b). The threshold of fusion of paired clicks has been compared to the threshold of fusion of paired flashes in an unpublished experiment. The difference between the two fusion thresholds was used as a measure of the relative dominance of one modality over the other. It was found that in normal subjects all two-click thresholds were lower than two-flash thresholds. Among the chronic schizophrenics who were tested there was in general a smaller difference between the two types of thresholds and among the very deteriorated patients there was a reversal of the normal findings: two-click thresholds were longer than two-flash thresholds. There was found to be a significant correlation between a rating of deterioration and the difference

between flash and click thresholds giving an impression as with the earlier studies of a greater disturbance in the auditory than the visual modality. It is an interesting parallel that simple RT to sound is significantly impaired under medication with LSD-25, while simple RT to light is not (Abramson, Jarvik, and Hirsch, 1955), and that under this drug normal subjects often display slowness, proverty of speech, and withdrawal (Rinkel, de Shon, Hyde, and Solomon, 1952), the symptoms that are displayed by the withdrawn deteriorated schizophrenic.

Zahn, Rosenthal, and Lawlor (1962) studied the GSR orienting reactions to visual and auditory stimulation in chronic schizophrenic and normal subjects. The basal log conductance during the series of stimuli was greater with visual than with auditory stimuli in schizophrenics while the reverse was the case with normals. This pattern was shown initially with the numbers of specific GSR's elicited but owing to the faster adaptation displayed by the normal subjects the pattern was not maintained after the first eight trials.

The same sort of alteration of the usual pattern of modality dominance is also shown in a study by Sutton, Hakerem, Zubin, and Portnoy (1961), where in addition to the effect of shifts in sensory modality which was the primary study of the experiment there was a general tendency for RT's to be faster to light than to sound with the chronic schizophrenic patients who were used as abnormal subjects.

This study and others on the same topic reviewed by Sutton and Zubin (1963) all show that with schizophrenics reactions to sound stimuli are impaired if the stimulus in the previous trial was a light, whereas reactions to light stimuli, however preceded, do not yield differences between patients and normals. This could be considered as further evidence for relative impairment of the auditory modality in schizophrenics.

Relative disorganization of the auditory modality in schizophrenics is a possible explanation for the results of the experiment of Fedio, Mirsky, Smith, and Parry (1961). These workers investigated the effect of EEG activation upon RT following the experiment of Lansing, Schwartz, and Lindsley (1959). Fedio et al. used the sound of a bell to evoke alpha blockade and presented a buzzer as a signal for reaction when the EEG alpha was blocked. They compared this condition with others where the EEG alpha was not blocked or was blocked spontaneously. It was found in contrast to the results from normal subjects that the speed of RT of schizophrenics did not increase when alpha was blocked. While not the interpretation put forward by the authors, disorganization of the auditory modality and of the means by which the auditory cortex is aroused is a possible explanation of the results.

To date the studies on the effectiveness of stimuli in the different

modalities and their relationship to schizophrenic pathology has not been given much attention. Work on this subject is not easy as other parameters such as intensity may produce swamping effects. However, attempts made, for instance, by Stevens (1955) to produce equivalent intensity scales for the visual and auditory modalities enable experimentation to be carried out.

B. INTENSITY

A pair of studies which belong partly in this and partly in the preceding section are those which consider the intensity of visual and auditory stimuli on the speed of reaction in schizophrenics (Venables and Tizard, 1956, 1958). In the first study eight intensities of stimulus light were used using a two-log unit range from 16 to 1500 foot-candles. It was found that in contrast to normal subjects, who showed a decrease in RT to an asymptotic value, the RT of chronic nonparanoid schizophrenics tended to increase after a fall to an optimal value at about 200 foot-candles, and became slower with increasing intensity. In contrast this "paradoxical" increase in RT was not shown with a wide range of auditory stimulus intensities up to 115 db with 1000-cps and 200-cps tones and white noise. With relation to the preceding section it is interesting to note that in comparing the effects of visual and auditory stimulus intensities it was evident that whatever the change of intensities the auditory RT tended to be slower than the visual for the nonparanoid chronic schizophrenics used in these studies.

More direct evidence on the effects of intensity of stimuli is given by the work of Shagass and Schwartz (1963a,b). Evoked cortical potentials were examined using averaging techniques to improve signal:noise ratio. Only the results of work on somato-sensory stimuli have been published so far, but in this it can be seen that in relatively acute schizophrenics there is a greater increase in evoked potential amplitude with increase in stimulus intensity than there is in normals. In this characteristic, schizophrenics are similar to other patients with the exception of dysthymic neurotics. This exception is of some theoretical interest as it has been postulated in some theories of schizophrenia that it is because the reactions of early schizophrenics are similar to those of anxiety states that the train of pathological behavior that eventuates in florid schizophrenia begins.

It should be noted that the portion of the evoked response that Shagass and Schwartz are dealing with in their studies is an early component in the evoked complex and there is some evidence (e.g., Allison, 1962) to indicate that it is a thalamo-cortical radiation response. The comparison of this neurophysiological evidence with behavioral evidence

is thus somewhat indirect. It will, however, be of direct interest to see whether a nonlinear relationship between visual stimulus intensity and the early component of evoked response appears in schizophrenics; if it does not, the explanation for the paradoxical RT findings must lie in later cortical functioning.

V. The State of the Patient

At the end of Section I,B two opposing viewpoints were mentioned concerning the relative levels of drive or arousal of acute and chronic schizophrenic states. Increasingly throughout the review of perceptual data the impression has gained ground that, if the findings on normal subjects (Section II,A) concerned with the motivational or arousal correlates of broadened or narrowed attention are applicable to the more long-term physiological disturbances involved in schizophrenic pathology, we shoud expect acute patients to be characterized by low, and chronic patients by high, levels of arousal or drive. There are, furthermore, suggestions that differences in physiological levels of activity may be found within the acute and chronic categories, depending upon the patient's subdiagnosis or state of withdrawal, or whether his past history indicates that he belongs to process or reactive, poor or good premorbid, categories.

The rather glib way in which the terms arousal, drive, and sympathetic or parasympathetic activity have been used so far in discussion may possibly be excused insofar as they are meant to be taken as shorthand summaries for concepts having wider implications. It is not the purpose of this chapter to present a detailed examination of these terms; nevertheless insofar as they have been invoked in explanation a word must be said about their usage. The interrelation between the cortex and such subcortical mechanisms as the hypothalamus and reticular formation is complex, but without doing too great violence to the finer points of detail it may be said that stimulation of both the posterior hypothalamus (which is responsible for sympathetic functions) and the dorsal region of the mesencephalic reticular formation causes cortical arousal signaled by such phenomena as desynchronization of the alpha rhythm. The functions of both these structures are blocked by barbituates and it is this principle which is involved in the sedation threshold test, the results of which are reviewed below. While in general it may be said that the anterior and posterior hypothalamus, as central structures mediating respectively parasympathetic and sympathetic function, operate in a reciprocal fashion so that activity in one tends to promote inhibition of the other, this is not universally the case. Acute

emotional disturbance, for instance, is accompanied by signs of both sympathetic and parasympathetic activity. Bearing this in mind in general we may say that conditions which produce activation of the central sympathetic mechanisms tend also to produce cortical arousal. On the other hand, drugs such as tranquilizers which shift the balance between the anterior and posterior hypothalamus to the parasympathetic side decrease the intensity of discharges from the hypothalamus and reticular formation to the cortex and lower the arousal level of the cortex.

The reaction of schizophrenic patients to drugs may allow us to draw conclusions about the patient's physiological state. Thus, if the action of a depressant drug is to improve the patient's condition, it is possible to make the inference that he was formerly in a state of hyperarousal.

One series of studies concerns the reaction of patients to such general depressants as sodium amytal. Lindemann (1932) was one of the first to show a temporary improvement in schizophrenic behavior after medication with sodium amytal. This was shown by increased contact with the environment and warmth of emotion. This work was repeated more recently, for instance, by Fulcher, Gallagher, and Pfeiffer (1957), who found that the number of "lucid intervals" among chronic schizophrenics increased under amobarbital. Stevens and Derbyshire (1958) reported remission of catatonic stupor under amobarbital and by taking EMG, EEG, and EKG recordings were able to show that remission from stupor was accompanied by a decrease in the level of cortical and autonomic activity. These studies all suggest a high degree of arousal in chronic deteriorated or catatonic patients. This idea is supported by the work on sedation threshold summarized by Shagass (1960). Amobarbital is given at a constant rate until the patient's speech is slurred or a point of inflection of the curve relating EEG fast activity to dose is reached. The amount of amobarbital required to reach this point is higher in chronic schizophrenics, other than those diagnosed as simple, than in normals. Patients diagnosed as acute schizophrenic, on the other hand, have the lowest of any sedation threshold reported. This low threshold in acute schizophrenics is confirmed in a study by Herrington and Claridge (1963). Anxiety states are also shown to have a high sedation threshold similar to that of chronic schizophrenics. The outcome of these studies is that there is evidence for high cerebral activity in chronic schizophrenia while the reverse is true of acute schizophrenic patients. This is based on the assumption that a patient in a state of high central activation requires a higher dose of barbiturate to eliminate that activity than one whose level

of activity is low. An alternative explanation may, however, be put forward in terms of a differential sensitivity to barbiturates. The study by Fulcher *et al.* (1957) mentioned above also used the drug arecoline and found that like amobarbital it produced "lucid intervals" in chronic schizophrenics. Arecoline is of interest because it has been shown to have pure parasympathetic (muscarinic) activity on the brain in comparison to amobarbital, whose action is more complex. Arecoline is one of the drugs like reserpine, pilocarpine, and physostigmine that inhibit conditioned avoidance in rats. It could be tentatively suggested that withdrawal in schizophrenics is a form of conditioned avoidance. In support of this it has been shown (Venables and Wing, 1962) that withdrawal in a nonparanoid chronic schizophrenic population is positively correlated with the level of skin potential shown by the subject. It has also been shown (Spain, 1963) that in a similar population the degree of eyelid conditioning (a paradigm for conditioned avoidance in humans?) is also positively related to the level of the subject's skin potential. Parasympatheticomimetic drugs may therefore act to produce remissions in deteriorated and withdrawn schizophrenics by decreasing conditioned avoidance and withdrawal. That avoidance conditioning is successful in schizophrenics is shown by the number of studies which give evidence of the greater effectiveness of aversive reinforcers in comparison with rewards in these patients (cf. Silverman, 1963).

Further support for the picture of higher activity in chronic schizophrenics is gained from the study of Williams (1953), who showed that during rest these patients showed higher basal levels of skin conductance, pulse rate, and respiration than normal. Similar results were also reported by Malmo, Shagass, and Smith (1951).

Within the chronic schizophrenic population, as has already been said, withdrawal has been shown to be related to the level of autonomic activity given by skin potential as an index and also to the level of cortical activity exemplified by two-flash fusion threshold (Venables and Wing, 1962). More recent unpublished data support this conclusion by a finding of a relationship between two-click threshold and withdrawal. These results, which show that the greater the degree of autonomic, visual cortical, or auditory cortical activation the greater is the withdrawal, apply only to nonparanoid schizophrenics and paranoid schizophrenics who are markedly incoherent. Standing apart from these are the coherent paranoid schizophrenics in whom there is a tendency (unpublished data) for withdrawal to be related to underarousal. In nonparanoid chronic schizophrenics two-flash threshold and skin potential as measures of cortical and autonomic activation are positively related. This pattern is

not found in normal subjects and coherent paranoid patients (Venables, 1963b). The atypical reaction of paranoid schizophrenics has been noted earlier in the review, and is confirmed by these latest results. Support for this view is given by such disparate studies as those of Stevenson, Derrick, Hobbs, and Metcalfe (1957), which suggests that paranoid schizophrenics differ from other schizophrenics in adrenocortical response and phosphate excretion, and by Wertheimer and Wertheimer (1955), who showed that the structure of capillary loops in all but paranoid schizophrenics was different from normals. This viewpoint is in agreement with that put forward by Shakow (1962), who says, "thus of the 58 measurements which we have made of a wide range of psychological functions on groups of normal, paranoid, and hebephrenic subjects, we found the paranoid to be nearer the normal in 31 instances and the hebephrenic nearer in only 7 instances. In 20 instances, however, the paranoid and hebephrenic fell on either side of the normal."

Among studies which are concerned with testing the difference between process and reactive schizophrenics one of the most convincing is that by Gromoll (1961), who, starting with the hypothesis that the process patients would be underaroused and the reactive patients overaroused, using percentage time alpha as his index, in fact found the complete reverse.

In this brief review of the physiological status of different classes of schizophrenic patients there has been the attempt only to consider the *level* of physiological activity. It would appear from perusal of the literature that many false conclusions have been drawn because of failure to distinguish level of ongoing activity from reactivity to stimuli. Attempts to infer one from the other tend to be inconclusive: Although, for instance, Wilder (1958) argues convincingly for the "law of initial value" by which there is an inverse linear relationship between the size of response and the level from which it starts, other workers—for instance, Silverman, Cohen, and Shmavonian (1959)—have argued for a curvilinear relationship between level and responses. It is by invoking the law of initial value, or indeed the initial part of the curvilinear relationship between level and response, that what appears at first sight to be a discrepancy between what is reported in the clinical literature and the state of the patient proposed in this section may be reconciled. Thus it is said that acute schizophrenic patients show large emotional reactions, anxiety, fear, anger, while the affect of the chronic patient tends to be "flat." It is against the proposed low level of activity of the acute schizophrenic that a large emotional response may be seen, while because of the high existing level of activity of the chronic patient only a small response may be evoked.

VI. Summary

The main thesis which has been put forward in this review is that, for the purpose of considering input dysfunction, schizophrenics cannot be taken as a whole. A distinction must be drawn between acute and chronic patients and, within this subdivision, attention must be paid to the patient's subdiagnosis and whether or not he may be designated as process or reactive. The interrelationships between the subdivisions have not become clear because little attempt has been made to sort them out. It is possible, for instance, that with the slow-onset, process-type schizophrenic we should not expect there to be an acute stage of the illness at all. On the other hand, as has been hinted at in the theory proposed by Fish (Section I,B), we should not perhaps ever think of a chronic stage of paranoid schizophrenia. If, as has become apparent from the experimental work reviewed, there is a marked distinction between the physiological states and hence the perceptual abilities of acute and chronic patients, far more experimental time can be profitably expanded on the analysis of the stage of transition from one to the other stage and the factors—physical, physiological, and social—which accelerate or retard the process. If we bear these points in mind, the experimental work which has been reviewed points to a general statement of the following kind. Chronic schizophrenic patients—and possibly included in this category are process patients—tend to be characterized by a state of restriction of the attentional field resulting from elevated states of sympathetic and cortical activation. Attention is restricted not only to the extent that peripheral sensory items contemporaneously present do not rise into consciousness, but also involved is the nonrecognition of items in memory which form part of the meaningful structure in which the present central item appears. Drugs and procedures which reduce arousal tend to improve the clinical status of these patients alongside (or even because of) an improvement in their perception of the external world and an increase in its meaningfulness resulting from increased breadth of attention.

In contrast to the chronic patient, the acute (and possibly the reactive and paranoid) patient is characterized by an inability to restrict the range of his attention so that he is flooded by sensory impressions from all quarters. Items of all kinds have equal importance, and the meaningfulness of the external world tends to be lost for the opposite reason to that which applies with the chronic patient. The figure-ground relationship which allows a picture to convey information is destroyed equally by making the picture nearly all figure or making it nearly all ground. The acute patient's broadened level of attention would appear

to arise from the low level of cortical activation or possibly the para-sympathetic imbalance which he displays. A wide variety of processes and skills seem to be affected by the attentional dysfunctions outlined, but nevertheless the impression should not be gained that what has been described is more than a segment of the total abnormality displayed by the schizophrenic patient.

References

Abramson, H. A., Jarvik, M. E., & Hirsch, M. W. L.S.D. 25 $\overline{X}$ Effect on RT to auditory and visual stimuli. *J. Psychol.*, 1955, **40**, 39-52.

Agnew, N., & Agnew, M. Drive level effects on tasks of narrow and broad attention. *Quart. J. exp. Psychol.*, 1963, **15**, 58-62.

Allison, T. Recovery functions of somatosensory evoked responses in man. *EEG clin. Neurophysiol.*, 1962, **14**, 331-343.

Ashman, G. R. Binary choice learning strategies in schizophrenics. Paper read at Am. Psychol. Assoc. meeting, Cincinnati, Ohio, 1959.

Bahrick, H. P., Fitts, P. M., & Rankin, R. E. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, **44**, 400-406.

Bharucha-Reid, R. P. The internal modulating system and stress: a neurophysiological model. *J. Gen. Psychol.*, 1962, **66**, 147-158.

Boardman, W. K., Aldrick, R. C., Reiner, M. L., & Goldstone, S. The effect of anchors upon apparent length. *J. gen. Psychol.*, 1959, **61**, 45-49.

Boardman, W. K., Goldstone, S., Reiner, M. L., & Fathauer, W. F. Anchor effects spatial judgment and schizophrenia. *J. abnorm. soc. Psychol.*, 1962, **65**, 273-276.

Boring, E. G. Size constancy and Emmerts Law. *Amer. J. Psychol.*, 1940, **53**, 293-295.

Bradley, R. B., & Elkes, J. Effect of some drugs on the electrical activity of the brain. *Brain*, 1957, **80**, 77-117.

Brown, G. W. Length of hospital stay and schizophrenia: a review of statistical studies. *Acta psychiat. et neurol. Scand.*, 1960, **35**, 414-430.

Callaway, E. The influence of amobarbital (amylobarbitone) and methamphetamine on the focus of attention. *J. ment. Sci.*, 1959, **105**, 382-392.

Callaway, E., & Band, R. I. Some psychopharmacological effects of atropine. *Arch. Neurol. Psychiat.*, 1958, **79**, 91-102.

Callaway, E., & Dembo, D. Narrowed attention: a psychological phenomenon that accompanies a certain physiological change. *Arch. Neurol. Psychiat.*, 1958, **79**, 74-90.

Callaway, E., & Stone, G. Re-evaluating the focus of attention. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960.

Callaway, E., & Thompson, S. V. Sympathetic activity and perception. *Psychosom. Med.*, 1953, **15**, 443-455.

Carson, R. C. Intralist similarity and verbal rote learning performance of schizophrenics and cortically damaged patients. *J. abn. soc. Psychol.*, 1958, **57**, 99-106.

Cleveland, S. E., Reitman, E. E., & Bentinck, C. Therapeutic effectiveness of sensory deprivation. *Arch. Gen. Psychiat.*, 1963, **8**, 455-460.

Conrad, K. *Die beginnende Schizophrenie.* Stuttgart, Germany: Thieme, 1958.

Davis, D. R. The disorganization of behavior in fatigue. *J. Neurol. Psychiat.*, 1946, **9**, 23-29. (a)

Davis, D. R. Neurotic predisposition and disorganization observed in experiments with the Cambridge cockpit. *J. Neurol. Psychiat.*, 1946, **9**, 119-124. (b)

Davis, J. M., McCourt, W. F., & Solomon, P. The effect of visual stimulation on hallucinations and other mental experiences during sensory deprivation. *Amer. J. Psychiat.*, 1959-1960, **116**, 889-892.

Drew, G. C. Mental fatigue. *Flying Personnel Research Committee*, 1940, Rep. No. 227.

Drew, G. C. The study of accidents. (Presidential address to the British Psychological Society), *Bull Brit. Psychol. Soc.*, 1963, **16** (52), 1-10.

Dunn, W. L. Visual discrimination of schizophrenic subjects as a function of stimulus meaning. *J. Pers.*, 1954, **23**, 48-64.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, **66**, 183-200.

Edwards, W. Apparent size of after images under conditions of reduction. *Amer. J. Psychol.*, 1953, **66**, 449-455.

Federn, P. *Ego psychology and the psychoses.* London: Imago, 1953.

Fedio, P., Mirsky, A. F., Smith, W. J., & Parry, D. Reaction time and EEG activation in normal and schizophrenic subjects. *EEG clin. Neurophysiol.*, 1961, **13**, 923-926.

Feinberg, I., & Garman, E. M. Studies of thought disorder in schizophrenia. *Arch. Gen. Psychiat.*, 1961, **4**, 191-201.

Fish, F. A neurophysiological theory of schizophrenia. *J. ment. Sci.*, 1961, **107**, 828-838.

Fulcher, J. H., Gallagher, W. J., & Pfeiffer, C. C. Comparative lucid intervals after amobarbital CO_2 and arecoline in chronic schizophrenics. *Arch. Neurol. Psychiat.*, 1957, **78**, 392-395.

Funkenstein, D. H., Greenblatt, M., & Solomon, H. C. Psychophysiological study of mentally ill patients. Part I: the status of the peripheral autonomic nervous system as determined by the reaction to epinephrine and mecholyl. *Amer. J. Psychiat.*, 1949, **106**, 18-28.

Garmezy, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1951, **20**, 253-276.

Gellhorn, E., & Miller, A. D. Methacholine and Noradrenaline tests. *Arch. Gen. Psychiat.*, 1961, **4**, 371-380.

Gibby, R. G., Adams, H. B., & Carrera, R. N. Therapeutic changes in psychiatric patients following partial sensory deprivation. *Arch. Gen. Psychiat.*, 1960, **3**, 33-42.

Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **50**, No. 2.

Greenberg, A. Directed and undirected learning in chronic schizophrenia. Unpublished Ph.D. Thesis, Columbia Univer., New York, 1953.

Gromoll, H. F. The process-reactive dimension of schizophrenia in relation to cortical activation and arousal. Unpublished Ph.D. Thesis, Univer. of Illinois, Urbana, 1961.

Hake, H. H., & Hyman, H. R. Perception of the statistical structure of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64-74.

Hamilton, V. Size constancy and cue responsiveness in psychosis. *Brit. J. Psychol.*, 1963, **54**, 25-39.

Harris, A. Sensory deprivation and schizophrenia. *J. ment. Sci.*, 1959, **105**, 235-236.

Hartman, A. M. The apparent size of after images in delusional and nondelusional schizophrenics. *Amer. J. Psychol.*, 1962, **75**, 587-595.

Hebb, D. O. *The organization of behavior.* London: Chapman & Hall, 1949.

Herrington, R. N., & Claridge, G. S. Personal communication, 1963.

Hoffer, A., & Osmond, H. F. Schizophrenia—an autonomic disease. *J. nerv. ment. Dis.*, 1955, **122**, 448-452.

Holway, A. H., & Boring, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21-37.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944, pp. 971-1032.

Huston, P. E., & Senf, R. Psychopathology of schizophrenia and depression. I: Effect of amytal and amphetamine sulphate on level and maintenance of attention. *Amer. J. Psychiat.*, 1952, **109**, 131-138.

Huston, P. E., & Singer, M. M. Effect of sodium amytal and amphetamine sulphate on mental set in schizophrenia. *Arch. Neurol. Psychiat.*, 1945, **53**, 365-369.

Kaplan, S. D. Autonomic visual regulation. *Psychiat. Res. Rep.*, 1960, **12**, 104-114.

Kinsey, A. C., Pomeroy, W. B., Martin, C. E., and Gebbard, P. H. *Sexual behavior of the human female*. Philadelphia, Pennsylvania: Saunders, 1953.

Klein, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation*. Lincoln: Univer. of Nebraska Press, 1954, pp. 224-274.

Knopf, I. J., & Fager, R. E. Differences in gradients of stimulus generalization as a function of psychiatric disorder. *J. abn. soc. Psychol.*, 1959, **59**, 73-76.

Kohn, H. Effects of variation of intensity of experimentally induced stress situations upon certain aspects of perception and performance. *J. genet. Psychol.*, 1954, **85**, 289-304.

Lansing, R. W., Schwartz, E., & Lindsley, D. B. Reaction time and EEG activation. *J. exp. Psychol.*, 1959, **58**, 1-10.

Leibowitz, H. W., & Pishkin, V. Perceptual size constancy in schizophrenia. *J. consult. Psychol.*, 1961, **25**, 196-199.

Lewinsohn, P. M., & Elwood, D. L. The role of contextual constraint in the learning of language samples in schizophrenia. *J. nerv. ment. Dis.*, 1961, **133**, 79-81.

Lhamon, W. T., & Goldstone, S. The time sense. Estimation of one second duration by schizophrenic patients. *Arch. Neurol. Psychiat.*, 1956, **76**, 625-629.

Lindemann, E. Psychological changes in normal and abnormal individuals under the influence of sodium amytal. *Amer. J. Psychiat.*, 1932, **11**, 1083-1091.

Lindsley, D. B. The reticular formation and perceptual discrimination. In H. H. Jasper, L. D. Proctor, R. S. Knighton, W. C. Noshay, & R. T. Costello (Eds.), *Reticular formation of the brain*. London: Churchill, 1958.

Lovinger, E. Perceptual contact with reality in schizophrenia. *J. abn. soc. Psychol.*, 1956, **52**, 87-91.

McGhie, A., & Chapman, J. S. Disorders of attention and perception in early schizophrenia. *Brit. J. med. Psychol.*, 1961, **34**, 103-116.

Malmo, R. B., Shagass, C., & Smith, A. A. Responsiveness in chronic schizophrenia. *J. Pers.*, 1951, **19**, 359-375.

Mednick, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. *J. abn. soc. Psychol.*, 1955, **51**, 536-542.

Mednick, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55**, 316-327.

Mednick, S. A., & DeVito, R. Associative competition and verbal learning in schizophrenia. Paper read at Eastern Psychological Assoc. Meeting, Philadelphia, 1958.

Miller, G. A., & Selfridge, J. Verbal context and recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176-185.

Nelson, S., & Caldwell, W. E. Perception of affective stimuli by normal and schizophrenic subjects in a depth perception task. *J. Gen. Psychol.*, 1962, **67**, 323-335.

Osgood, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. of Nebraska Press, 1957.

Payne, R. W. Cognitive abnormalities. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961.

Payne, R. W. An object classification test as a measure of overinclusive thinking in schizophrenic patients. *Brit. J. soc. clin. Psychol.*, 1962, 1, 213-221.

Payne, R. W., & Friedlander, D. A short battery of simple tests for measuring overinclusive thinking. *J. ment. Sci.*, 1962, 108, 362-367.

Payne, R. W., Matussek, P., & George, E. I. An experimental study of schizophrenic thought disorder. *J. ment. Sci.*, 1959, 105, 627-652.

Pearl, D. Language processing ability of process and reactive schizophrenics. *J. Psychol.*, 1963, 55, 419-425.

Raush, H. L. Perceptual constancy in schizophrenia. *J. Pers.*, 1952, 21, 176-187.

Rinkel, M., DeShon, H. J., Hyde, R. W., & Solomon, H. Experimental schizophrenia like symptoms. *Amer. J. Psychiat.*, 1952, 108, 572-578.

Rodnick, E. H., & Shakow, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, 97, 214-225.

Rosenthal, D., Lawlor, W. G., Zahn, T. P., & Shakow, D. The relationship of some aspects of mental set to degrees of schizophrenic disorganization. *J. Pers.*, 1960, 28, 26-38.

Rosenzweig, N. A mechanism in schizophrenia. *Arch. Neurol. Psychiat.*, 1955, 74, 544-555.

Rosenzweig, N. Sensory deprivation and schizophrenia: some clinical and theoretical similarities. *Amer. J. Psychiat.*, 1959-1960, 116, 326-329.

Rump, E. E. A note on "Distance constancy in schizophrenic patients." *J. ment. Sci.*, 1961, 107, 48-51. (a)

Rump, E. E. The relationship between perceived size and perceived distance. *Brit. J. Psychol.*, 1961, 52, 111-124. (b)

Salzinger, K. Shift on judgment of weights as a function of anchoring stimuli and instructions in early schizophrenics and normals. *J. abn. soc. Psychol.*, 1957, 55, 43-49.

Samuels, I. Reticular mechanisms and behavior. *Psychol. Bull.*, 1959, 56, 1-25.

Schwartz, M., & Shagass, C. Reticular modification of somatosensory cortical recovery function. *EEG clin. Neurophysiol.*, 1963, 15, 265-271.

Shagass, C. Drug thresholds as indicators of personality and affect. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960.

Shagass, C., & Schwartz, M. Cerebral responsiveness in psychiatric patients. *Arch. Gen. Psychiat.*, 1963, 8, 177-189. (a)

Shagass, C., & Schwartz, M. Psychiatric disorder and deviant cerebral responsiveness to sensory stimulation. *Recent advances in biol. Psychiat.*, 1963, 5, 321-330. (b)

Shakow, D. Segmental set. *Arch. Gen. Psychiat.*, 1962, 6, 1-17.

Shakow, D. Psychological deficit in schizophrenia. *Behavioral Sci.*, 1963, 8, 275-305.

Shannon, C. E. Prediction and entropy of printed English. *Bell Syst. Tech. J.*, 1951, 30, 50-64.

Silverman, A. J., Cohen, S. I., & Shmavonian, B. M. Investigation of psychophysiologic relationships with skin resistance measures. *J. Psychosom. Res.*, 1959, 4, 65-87.

Silverman, J. Psychological deficit reduction in schizophrenia through response-contingent noxious re-inforcement. *Psychol. Rep.*, 1963, 13, 187-210.

Smith, S., Thakurdas, H., & Lawes, T. G. G. Perceptual isolation and schizophrenia. *J. ment. Sci.*, 1961, 107, 839-844.

Snygg, D., & Combs, A. W. *Individual behavior*. New York: Harper, 1949.

Spain, B. Personal communication, 1963.

Steriade, M., & Demetrescu, M. Reticular facilitation of responses to acoustic stimuli. *EEG clin. Neurophysiol.*, 1962, 14, 21-36.

Stevens, J. M., & Derbyshire, A. J. Shifts along the alert-repose continuum during remission of catatonic stupor with amobarbital. *Psychosom. Med.*, 1958, 20, 99-107.

Stevens, S. S. Decibels of light and sound. *Physics today*, 1955, 8, 12-17.

Stevenson, J. A. F., Derrick, J. B., Hobbs, G. E., & Metcalfe, E. V. Adrenocortical response and phosphate excretion in schizophrenia. *Arch. Neurol. Psychiat.*, 1957, 78, 312-320.

Stroop, J. R. Studies of interference in serial verbal reactions. *J. exp. Psychol.*, 1935, 18, 643-661.

Sutton, S., & Zubin, J. Effect of sequence on reaction time in schizophrenia. Paper presented to International Colloquium on Biological Bases of Age Changes in the Speed of Behavior. Cambridge, England, 1963.

Sutton, S., Hakarem, G., Zubin, J., & Portnoy, M. The effect of shift of sensory modality on serial reaction time: a comparison of schizophrenics and normals. *Amer. J. Psychol.*, 1961, 74, 224-232.

Teichner, W. H. Recent studies of simple reaction time. *Psychol. Bull.*, 1954, 51, 128-149.

Thouless, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, 22, 216-241.

Topping, G. G., & O'Connor, N. The response of chronic schizophrenics to incentives. *Brit. J. med. Psychol.*, 1960, 33, 211-214.

Venables, P. H. A short scale for rating "activity-withdrawal" in schizophrenics. *J. ment. Sci.*, 1957, 103, 197-199.

Venables, P. H. The effect of auditory and visual stimulation on the skin potential response of schizophrenics. *Brain*, 1960, 83, 77-92.

Venables, P. H. Selectivity of attention, withdrawal, and cortical activation. *Arch. gen. Psychiat.*, 1963, 9, 74-78. (a)

Venables, P. H. The relationship between level of skin potential and fusion of paired light flashes in schizophrenic and normal subjects. *J. Psychiat. Res.*, 1963, 1, 279-287. (b)

Venables, P. H., & O'Connor, N. Reaction times to auditory and visual stimulation in schizophrenic and normal subjects. *Quart. J. exp. Psychol.*, 1959, 11, 175-179.

Venables, P. H., & Tizard, J. Paradoxical effects in the reaction time of schizophrenics. *J. abn. soc. Psychol.*, 1956, 53, 220-224.

Venables, P. H., & Tizard, J. The effect of auditory stimulus intensity on the reaction time of schizophrenics. *J. ment. Sci.*, 1958, 104, 1160-1164.

Venables, P. H., & Wing, J. K. Level of arousal and the subclassification of schizophrenia. *Arch. gen. Psychiat.*, 1962, 7, 114-119.

Weckowicz, T. E. Size constancy in schizophrenic patients. *J. ment. Sci.*, 1957, 103, 475-486.

Weckowicz, T. E. Autonomic activity as measured by the mecholyl test and size constancy in schizophrenic patients. *Psychosom. Med.*, 1958, 20, 66-71.

Weckowicz, T. E., & Blewett, D. B. Size constancy and abstract thinking in schizophrenic patients. *J. ment. Sci.*, 1959, 105, 909-934.

Weckowicz, T. E., Sommer, R., & Hall, R. Distance constancy in schizophrenic patients. *J. ment. Sci.*, 1958, 104, 1174-1182.

Weinstein, A. D., Goldstone, S., & Boardman, W. K. The effect of recent and remote frames of reference on temporal judgments of schizophrenic patients. *J. abn. soc. Psychol.*, 1958, 57, 241-243.

Wells, F. L., & Kelley, C. M. The simple reaction in psychosis. *Amer. J. Psychiat.*, 1922, 2, 53-59.

Wertheimer, N., & Wertheimer, M. Capillary structure: its relation to psychiatric diagnosis and morphology. *J. nerv. ment. Dis.*, 1955, 122, 14-27.

Wilder, J. Modern psychophysiology and the law of the initial value. *Amer. J. Psychother.*, 1958, 12, 199-221.

Williams, M. Psychophysiological responsiveness to psychological stress in early chronic schizophrenic reactions. *Psychosom. Med.*, 1953, 15, 456-461.

Winer, H. R. Incidental learning in schizophrenics. Unpublished Ph.D. Thesis, Purdue Univer., Lafayette, Indiana, 1954.

Zahn, T. P., Rosenthal, D., & Lawlor, W. G. G.S.R. orienting reactions to visual and auditory stimuli in chronic schizophrenic and normal subjects. Paper read to Society for Psychophysiological Research, Denver, Colorado, 1962.

Zahn, T. P., Rosenthal, D., & Shakow, D. Effects of irregular preparatory intervals on reaction time in schizophrenia. *J. abn. soc. Psychol.*, 1963, 67, 44-52.

A THEORY OF VERBAL BEHAVIOR IN SCHIZOPHRENIA[1]

Loren J. Chapman, Jean P. Chapman, and Glenn A. Miller

SOUTHERN ILLINOIS UNIVERSITY, CARBONDALE, ILLINOIS

I. Theoretical Position

A. GENERAL PRINCIPLE

Schizophrenics often misinterpret and misuse words in common discourse. Although this observation is commonplace, and despite the fact that several diverse clinical formulations of the phenomenon have gained popularity, very little has been done in the way of systematic exploration of the nature of these errors. The present paper reports three experiments that are directed toward testing predictions from one formulation which may account for many of these errors.

The theory which will be presented here was stimulated in part by earlier findings which indicate a close similarity between schizophrenic disorder of thought and normal error tendencies. There is a large literature which discusses the similarity between schizophrenic thought dis-

order and error tendencies of normal subjects under one or another of various special conditions which induce errors. These special conditions include sleep, psychotomimetic drugs, sensory deprivation, fatigue, relaxed attention, emotional excitement, and oxygen deprivation. However, the findings of a number of studies have indicated that a special condition is unnecessary. In a series of studies designed to measure clinically prominent features of the disorder, it was repeatedly found that the kinds of errors that were hypothesized to be features of schizophrenic disorder were also found in normal control subjects, although in reduced number. This was true of (1) associative intrusions into schizophrenic conceptual performance (Chapman, 1958), (2) the interpretation of common concepts in an excessively broad manner so as to include incorrect items which are similar to correct ones (Chapman and Taylor, 1957; Chapman, 1961), (3) the confusion of figurative and literal usages of common words and phrases (Chapman, 1960), and (4) the tendency to solve formal syllogisms by concluding an identity of objects which share a common quality (Chapman and Chapman, 1959; Gottesman and Chapman, 1960).

 There are reasons for believing that these findings reflect an accentuated expression, in schizophrenia, of normal response biases. It is as if the response biases are released and expressed more freely. A response bias may be defined as a predisposition toward making a particular one of the various possible kinds of responses that one might make to a given stimulus (Underwood, 1952). It is called a "bias" because psychologists often define it by responses which are incorrect or inappropriate in the context in which they appear, but it may produce correct answers in those situations in which a biased response and the correct response coincide. [Campbell (1959) has catalogued a number of these biases.]

 For both normal and schizophrenic persons, response biases act more strongly in situations in which other stimuli are not available as cues which limit the variety of responses that are appropriate. For example, one strong normal response bias is to respond to a word with its associate. (An associate here means that kind of response such as is given on a word association test.) The reader may observe for himself just how strong the associative bias is by trying to think of a list of other words that are *not* associatively related to a single stimulus word—he will probably find that one strong associative response after another comes to mind. In normal discourse the range of appropriate responses to a word is defined by the context in which the word appears. This context supplies a multitude of cues which guide the selection of the overt response to the word. This usually tends to eliminate associative responses. Nevertheless, associates occasionally intrude as inappropriate responses in normal speech. However, they occur far more freely if the contextual cues are

reduced. For example, they occur in the word association test itself in which such contextual cues are minimal since the instruction for the test specifies that the S is to say the first word that comes to mind when he hears a single stimulus word. Associates are also abundant in the free association of normal persons, as on the psychoanalytic couch.

It appears clinically that many schizophrenics free-associate much of the time in place of more goal-directed discourse. Similarly, Chapman (1958) found that normal associates intrude as incorrect responses in schizophrenic conceptual performance more than in normal performance. The observation that for normal persons the associative bias is expressed more freely when constraining contextual cues are reduced suggests the possibility that a reason for the schizophrenic's excessive yielding to this normal response bias may be a lesser responsiveness to contextual cues. These are the cues which, for normal persons, guide the selection of appropriate responses in place of biased ones. For example, Bleuler's (1950, p. 26) patient who, in describing a walk with her family began with enumerating the members of her family, listed "Father, son" and then added, "and the Holy Ghost." The context of the sentence would have indicated to a normal person that the response "and the Holy Ghost" was inappropriate even though it would have been an appropriate response to "Father, son" in some other contexts. The utterance then may be seen as reflecting a lesser responsiveness to the contextual cues.

These considerations indicate that many features of the schizophrenic disorder of thought consist of an accentuated expression of those overt responses toward which normal subjects are biased. It seems likely that an elaboration of this principle will explain the schizophrenic misinterpretation and misuse of the meanings of words. This elaboration was originally suggested by spontaneous comments that some schizophrenic and normal Ss made during testing when presented with items which asked whether two words have the same meaning. One item asked if *bicycle* means the same as *wagon*. The schizophrenics often said things like, "You can ride on a bicycle, and you can ride in a wagon, so they mean the same." Normal Ss would say something like, "You can ride on both a bicycle and in a wagon, but a bicycle has two wheels and a wagon has four, so they don't really mean the same." These responses suggested that schizophrenics do not weigh simultaneously the several different aspects of meaning in order to answer appropriately the question at hand, but instead answer by using a more limited number of aspects of the meaning. Moreover, the aspects of meaning which schizophrenics use to excess appear to be those which are more prominent for normal persons. The tendency to interpret a word in accordance with its most prominent aspect of meaning is a response bias in the same sense that

the tendency to respond to a word with associates is a response bias. This may be illustrated most readily with a double-meaning word. If a normal person is asked for the meaning of the word *grain*, he is most likely to define it in terms of various food plants such as wheat or oats. However, if he encounters the word *grain* in the context of a sentence which contains appropriate cues, he will instead interpret it to mean the lines or markings in wood.

As will be seen below (Section I, C) one may infer from schizophrenics' misinterpretation of words a possible explanation of why they fail to respond to the contextual cues which, for normal persons, restrain the expression of response biases. But first we will present further considerations and data on this phenomenon of schizophrenic misinterpretations of the meanings of words.

Actually the phrase, "the meaning of a word," is somewhat inexact, since "meaning" does not reside in the spoken or written word itself, or its physical properties, but instead exists only as a response within the person who uses or encounters the word. It is probably clearer, therefore, in discussing research and theory concerning meaning, to speak of "meaning responses" instead of "aspects of meaning" of a word. A meaning response, according to this usage, is a hypothetical internal event which mediates a person's overt behavioral response to a word.

Several previous writers have used the term "meaning response" and some of them have offered measures designed to define it operationally. Some of these have been concerned with "affective" or "connotative" meaning as measured by the semantic differential (Osgood, Suci, and Tannenbaum, 1957). For example, if one is investigating the meaning of *dog*, the semantic differential would produce measures of how good or bad, how weak or strong, and how quick or slow the concept *dog* is felt to be. Other investigators have been concerned with association as a measure of meaning (Bousfield, Whitmarsh, and Danick, 1958; Deese, 1962). Associative measures of meaning might indicate that the aspects of meaning of *dog* include "cat."

These two approaches may yield valuable information about important varieties of "meaning," and the theory of this paper yields predictions which might be tested in terms of them. However, the present paper is primarily concerned with another variety of "meaning" which is usually called "denotative meaning." Denotative meaning is difficult to define, but it is what the layman usually means by the meaning of a word. It is the representative of the referent of the word, the objects or events to which it refers. When the referent is an object, the denotative meaning is in large part a description of the object. For example, the denotative meaning of *dog* includes "is an animal," "has four legs,"

etc. Meaning responses are measured by obtaining statements of meaning from judges. The judges are instructed to "tell what the thing named is, or what it is like," or to list "aspects of meaning" for the word. Evidence concerning relative strength of meaning responses is obtained either from the order in which the statements of meaning are listed, or by asking judges to rank order them. In this presentation, we are limiting the term "meaning responses" to internal events which represent the referents of a word, and "statements of meaning" to the overt responses that judges give to questions concerning the aspects of meaning of a word.

Stated in terms of meaning responses, the theory states that schizophrenics' misinterpretations of the meanings of words arise in part from mediation of overt responses to words by their strongest meaning responses with a relative neglect of their weaker meaning responses, while the interpretation of words by normal persons reflects the use of the weaker as well as the stronger meaning responses. It is assumed that meaning responses themselves are similar for schizophrenics and normals. Thus predictions of group tendencies in schizophrenic thinking are made using the average strength of the various statements of meaning offered by a group of normal judges. Variation among schizophrenics in their thinking disorder is predicted using variation in the statements of meaning by normal judges.

B. Previous Work

Chapman and Chapman (1964) found support for this theory from a study of the characteristics of words which schizophrenics to an excessive extent regard as synonymous. The tendency of schizophrenics to regard some words of only moderate similarity as synonymous is a fairly striking feature of schizophrenic thought disorder, and it is manifested in their tendency to substitute words for one another in ways that normal persons do not. These words often have a similarity that normal persons recognize, but do not regard as sufficient to define the words as synonymous. The present theory states that schizophrenics tend to interpret words in accordance with the strongest meaning responses, neglecting weaker meaning responses. Therefore it predicts that the schizophrenic tendency to treat words of similar meaning as synonymous will occur more often on pairs for which the same meaning response is strongest for both of the two words. For example, 31 of 52 normal judges listed "is an animal" as the strongest statement of meaning for both *pig* and for *dog*. It follows from the theory that schizophrenics, more than normal subjects, should regard *pig* and *dog* as synonymous.

A contrasting word pair is *brassiere–tee shirt*. The five statements

of meaning for *brassiere*, listed in order of rated strength, were "supports breasts," "is an undergarment," "clothing," "used by women," and "made of cloth." The five statements of meaning for *tee shirt* were "is clothing," "is an undergarment," "used by fellows (or men)," "made of cloth," and "is white." Although these two words have in common three of their five statements of meaning, only 13 of 52 normal judges ranked the same statement of meaning as first in strength for both of the two words. The prediction from the theory, therefore, was that schizophrenics would less often deviate from normal control Ss by calling pairs like *brassiere* and *tee shirt* synonymous than by calling pairs like *pig* and *dog* synonymous.

In order to test this formulation 38 pairs of words were dichotomized by the measure of the number of judges who ranked the same statement of meaning in first position of strength for both members of a word pair. In addition, ratings of degree of similarity between pairs of words were obtained in separate judgments. It was found that for words of the range of similarity represented in the study, the measure of shared high-strength meaning responses had little relationship to degree of similarity. Similarity may either consist of several low-strength shared meaning responses or of a single high-strength meaning response. For example, *pig–dog* and *brassiere–tee shirt* were close on rated similarity despite their disparity on shared high-strength meaning responses. Two 19-item sets of word pairs were obtained which differed on whether their shared meaning responses were high strength or low strength.

The tendency to respond to word pairs as if they were synonymous was measured by items of the following format:

7. "Pig" means the same as
   A. stocking
   B. dog
   C. neither of the above.

The schizophrenics were found to exceed the normal subjects on the difference between the number of pairs of the two sets which were called the same ($p < .001$). The schizophrenics judged as synonymous more pairs of high-strength shared meaning response than low-strength pairs ($p < .03$), while the normals showed qualitatively the opposite pattern. These differences were found to be independent of the rated similarity between the pairs of words. These findings indicate that schizophrenics are more likely to show an abnormal propensity for regarding two words as synonymous if the words have the same strongest meaning response in common. The findings support the contention that schizophrenics' overt responses are mediated to an excess by the stronger normal meaning responses, with a relative neglect of weaker meaning responses.

C. PREDICTIONS

We have restricted the definition of meaning responses to internal events which represent the denotative referents of a word. Meaning responses to the specific words which are being interpreted are only one variety of mediating response. As mentioned previously other external stimuli from the context in which the word appears also have an influence on which of the several meaning responses to a word are used to mediate the overt response. The most important of these contextual cues are other words. (However, they also include such stimuli as vocal qualities and gestures.) Presumably, these contextual cues also act through internal mediating responses. As discussed above (Section 1, A), it appears that schizophrenics often fail to respond to the contextual cues which have a restraining influence on the expression of response biases for normal persons. The present findings and theory indicate a possible explanation for this. It seems likely that the schizophrenic excessive use of the strongest meaning responses also holds for other types of mediating responses, such as those aroused by contextual cues, instructions, etc. It was found that schizophrenics tend to mediate their responses to a word only by the strongest normal meaning response, neglecting the weaker. It would then appear reasonable to assume that there is a general propensity to use the strongest normal mediating response with a neglect of the weaker whether the mediating response is a meaning response or is instead another kind of mediating response. The mediating response to each such contextual cue which guides the selection of overt responses may frequently be weak in comparison to the mediating response which evokes the biased overt response. For example, if a person is asked to interpret the meaning of the word *rare* in the statement, "Robert says he likes rare meat," the word *meat* provides a weak contextual cue which for normal persons indicates that the correct meaning is "partially cooked" rather than "uncommon," which is a stronger meaning for rare when contextual cues are absent. The theory predicts that it is in cases such as this, in which the contextual cues are weak, that schizophrenics fail to use these cues. Instead they interpret words in accordance with the meaning responses to the words which are strongest when the words are encountered out of context. The frequent failure of schizophrenics to use contextual cues would then be seen as one manifestation of the tendency to guide behavior by the strongest normal mediating response. It follows that in situations in which the contextual cues are stronger schizophrenics will use them more often and their performance will approach that of normals. (Experiment III, below, will provide some evidence on this.)

Thus restated, the theory postulates that schizophrenics' misinter-

pretations of the meanings of words arise in part from mediation of behavior by the strongest normal mediating responses with a relative neglect of weaker mediating responses. It should be noted that this theory does not attempt to account for the deviant performance of schizophrenics on verbal tasks by postulating a qualitative difference in the mediating responses. Instead the theory assumes that the mediating responses of schizophrenics show a marked similarity to those of normal persons. The theory states that much schizophrenic misinterpretation of words reflects an excessive reliance on the mediating responses which are strongest for normal persons. Strength, as referred to in this paper, always applies to the degree of relationship between a stimulus and a response, but following the convention of behavior theorists, we will occasionally refer to it in terms of strength of cues or strength of responses. In all cases, however, it is the strength of the "bond" between stimulus and response that is implied.

The assumption of the similarity of schizophrenics and normal persons in the content and relative strength of meaning responses was investigated in a pilot study. Statements of meaning to 20 common words were obtained from a group of 20 chronic schizophrenics and a group of 22 normal control subjects of similar age and education. Of these 20 words, 15 had more than one meaning and 5 had only a single meaning. It was found that for 19 of these 20 words, the statement of meaning which was most frequently given by the normal group was also the statement of meaning most frequently given by the schizophrenics. As a further check on the comparability of the two groups, each subject was given a score, for the 15 multimeaning words, of the number of first position statements of meanings that coincided with the meaning most commonly given by the normal group as a whole. The mean of these scores was 13.5 for the schizophrenics and 13.7 for the normals. The difference was not significant ($t = .46$; $p = .64$). These results lend strong support to the assumption that schizophrenics and normals have a similar hierarchy of meaning responses to common words.

It should also be noted that the theory offers no exact specifications of the nature of the interaction of schizophrenia with mediating responses. The hypothesized excess in schizophrenia of overt responses mediated by the strongest mediating responses could be accounted for by several alternative formulations concerning the corresponding internal events. It could be explained by a strengthening of the strong mediating responses, or by a weakening of the weak mediating responses, or by a multiplication of the strength of both strong and weak mediating responses, or by a selective failure to respond to weak mediating responses, or by some selective inhibitory process. There is no evidence at

present for choosing among these alternative explanations. In addition, even if one could arrive at such a specific formulation in terms of mediating responses, one could not assume an isomorphic representation of this pattern in overt behavior.

One of the predictions from the theory in terms of data is that schizophrenics, much more than normals, will interpret words in accordance with the strongest normal mediating responses whether or not these are appropriate. This means that their errors will be more numerous when the correct overt response is mediated by a weak mediating response, as measured by normal group judgments, and these errors will consist of interpreting the word in accordance with a strong normal mediating response. When the correct overt response is mediated by the strongest normal mediating response, schizophrenic accuracy will approach that of normals. Although this differential error propensity will be found for both schizophrenics and normals, it will be greater for schizophrenics. Moreover, these effects will be greater when normal judges show greater agreement as to which is the strongest mediating response.

II. Experimental Investigations

A. METHOD AND SUBJECTS

The present paper reports three studies which tested these predictions. All three studies used, as a research tool, words which have more than one meaning in ordinary usage. For example, a pen means both "a writing implement" and "a fenced enclosure." The first of these is by far the stronger of the two meanings, as shown by our measures of strength of meaning response. While the theory is intended to apply to the use of all words by schizophrenics, its implications can be conveniently tested with multimeaning words.

These three experiments were not performed in the order in which they are presented, and for all of them the data, except the measures of strength of meaning responses, were gathered prior to the formulation of the theory.

1. Schizophrenic Subjects

The patients in all three studies were long-term chronic schizophrenics who were not receiving drug therapy. They were all native-born white male patients whose last admission had been no more recent than 6 months, and most of whom had spent many years in the hospital. No patient was included who showed, either by interview or by examination of the hospital records, indications of brain damage or mental deficiency,

or who had received electric shock or insulin shock therapy in the previous 3 months, or "tranquilizing" drugs for the previous 4 weeks. The patients used in Experiments I and III were from a group of patients at Chicago State Hospital who were being maintained without drugs for purposes of other research. The patients in Experiment II were at Kentucky State Hospital in Danville.

For several years now a major obstacle to conducting meaningful research on schizophrenic thinking has been that many hospitalized schizophrenics are kept on maintenance dosages of one or another of the phenothiazines, commonly chlorpromazine. Since the phenothiazines markedly affect clinical symptoms, they might also be expected to affect performance on any research task which reflects psychopathology. A number of studies have tested for drug effects on psychological measures and most have found them. Some of these have found the phenothiazine improves performance on a given measure, while other studies have found that it impairs performance. Chapman and Knowles (1964) have shown recently that phenothiazines produce an increase in some kinds of cognitive errors, and a decrease in others. They found that the drugs exacerbate errors that might be described as unsystematic, careless, or random, but reduce those that reflect a misinterpretation of common concepts in an overly broad manner. The latter kind of error is a striking feature of the disorder of thought in schizophrenia and is probably highly akin to some of the kinds of errors that were studied in the present research. It seems clear from these results that patients receiving the phenothiazines are an unsuitable group for the study of schizophrenic thought disorder.

It is also undesirable to use those patients that the hospital personnel find can be conveniently maintained without drugs without a worsening of the symptoms. Chapman (1963) has shown that selection of such patients can bias the results of research concerned with intellectual functioning in schizophrenia since patients whose symptoms worsen with the withdrawal of phenothiazine tend to be those with the more severe cognitive disorder. The present investigators were fortunate in being able to obtain for all of these studies chronic schizophrenics who were maintained off drugs regardless of their symptoms. Thus a major source of potential bias in the data was avoided.

Another source of possible bias that could not be completely avoided was the exclusion from the study of patients who were unable or unwilling to complete testing. Although this bias is, of course, unavoidable, it was somewhat reduced in the present studies by the use of tests of very simple format, and by extended efforts to develop rapport with the patients over a period of weeks.

2. Normal Subjects

The normal control Ss in Experiments I and III were City of Chicago firemen. Those in Experiment II were hired from among maintenance workers at the University of Kentucky, and from among unemployed workers at the Kentucky Employment Service. Although the Ss from the Employment Service were currently unemployed, no one

TABLE I

MEAN AND VARIANCE OF AGE, EDUCATION, AND STANFORD–BINET VOCABULARY SCORE FOR TOTAL GROUPS AND SUBGROUPS MATCHED ON VOCABULARY

	N	Age		Education		Vocabulary	
		$\overline{X}$	s^2	$\overline{X}$	s^2	$\overline{X}$	s^2
Experiment I							
Total groups							
Normal	27	43.2	36.6	10.6	5.2	21.7	29.2
Schizophrenic	25	44.4	47.6	10.7	4.0	20.3a	52.2a
Experiment II							
Total groups							
Normal	22	40.4	93.3	10.7	9.8	23.2	20.6
Schizophrenic	49	47.3	67.6	9.8	11.6	16.4b	35.3b
Matched subgroups							
Normal	15	39.6	100.0	10.9	11.2	22.1	21.2
Schizophrenic	15	51.2	33.2	11.6	16.4	22.1	20.8
Experiment III							
Total groups							
Normal	25	38.7	58.4	11.0	2.5	23.9c	29.6c
Schizophrenic	27	45.3	16.5	10.0	5.7	19.8d	58.5d
Matched subgroups							
Normal	20	38.4	48.4	11.0	2.4	22.8	20.7
Schizophrenic	20	46.4	11.3	9.8	5.6	22.9	21.7

a Information on only 21 subjects.
b Information on only 44 subjects.
c Information on only 22 subjects.
d Information on only 25 subjects.

was included who had not held a job for at least 6 months during the previous year. A number of them were currently out of work during the winter months, but normally worked steadily in the building trades during warmer weather. The normal control Ss of Experiment II were paid either $5.00 or $10.00 to participate in a number of experiments including the present ones. The patients for whom they were controls were not paid cash but were given a package of cigarettes or a package of chewing tobacco for each test taken. There were no such payments in Experiments I and III.

Table I shows the mean age, education, and vocabulary score for the schizophrenic and control groups for each of the three studies.

B. EXPERIMENT I. MISINTERPRETATION IN CONTEXT

Experiment I was used to test predictions concerning the misinterpretation of double-meaning words which are presented in context.

There are many common English words which have more than one meaning in everyday usage. For example, the word *rare* may mean either "uncommon" or "partially cooked" and a person who receives a communication which includes the word *rare* must decide which meaning is intended on the basis of cues obtained from the context in which the word appears. These cues may be strong or weak or—to state it in terms of mediating responses—they may evoke strong or weak mediating responses. As will be seen below, the present study was concerned with words presented in sentences which contain only weak cues.

When a multimeaning word is encountered alone, i.e., outside the context of other words, the various meanings differ in salience, i.e., the meaning responses which convey the various meanings differ in strength.

If schizophrenics have an inadequate response to weak cues provided by the context in which a multimeaning word appears, their interpretation of multimeaning words in sentences with such weak cues should show an excessive reliance on those meaning responses which are strongest when the words are encountered out of context. Hence, when the correct interpretation of a word in a sentence is consistent with a relatively weak meaning response, schizophrenics should misinterpret the word, and their misinterpretation should be one mediated by a stronger meaning response. This kind of error should exceed the opposite error of making an inappropriate interpretation mediated by a weaker meaning response when a stronger one is appropriate.

Method. Items were constructed around two common meanings of multimeaning words in the following format.

21. When the farmer bought a herd of cattle, he needed a new pen.
This means:
   A. He needed a new writing implement.
   B. He needed a new fenced enclosure.
   C. He needed a new pick-up truck.

Here the correct answer is intended to be alternative "B," although the answer in alternative "A" is mediated by the stronger meaning response to *pen*. This item was paralleled by the following item.

40. The professor loaned his pen to Barbara.
This means:
   A. He loaned her a pick-up truck.
   B. He loaned her a writing implement.
   C. He loaned her a fenced enclosure.

Here the correct answer was intended to be alternative "B," and alternative "C" contains a possible misinterpretation mediated by the weaker meaning response. It should be noted in item 21 above that the cues from the context are related somewhat weakly to the meaning responses to *pen*. For example, stronger cues would be given for the weaker meaning of *pen* by the statement, "The farmer put his cattle in the pen." All of the items of the present study were constructed so as to make the cue strength as weak as possible but still of sufficient strength to indicate the correct choice to normal persons.

It is seen that in addition to the alternatives representing the correct and the incorrect meanings of the word there is a third alternative in each item which is unrelated to the word *pen*. This alternative, called here "the unidentified error alternative," was included as a measure of "random" responses, i.e., some Ss might mark either of the other two alternatives for reasons other than their understanding of the statement that they are asked to interpret. For example, schizophrenics might be uncooperative or have difficulty taking the test and so might mark randomly. Such random marking should result as often in the marking of the unidentified error alternative as in marking of the other two alternatives. Therefore, on items measuring misinterpretations mediated by the stronger meaning responses, the number of unidentified error alternatives for each S was subtracted from the number of markings of the alternatives representing the stronger meaning responses, to obtain a corrected measure of the predicted error tendency. A similar correction was used with items measuring misinterpretations mediated by the weaker meaning responses. These corrected scores were used in all analyses of the data.

At test of 38 items was constructed in the format of the above items. It consisted of one item for each of two meanings of 19 double- or multi-meaning words. Table II lists the 19 words together with phrases which indicate the two meanings represented in the alternatives.

Evidence concerning the relative strength of the two meanings of each multimeaning word was obtained by presenting a list of the words to a group of normal judges (39 students in introductory psychology), with the instruction: "What is the first aspect of meaning you think of for each of the following words, what is the second, and what is the third?" The resultant statements of meaning were weighted in the following manner. When a statement of meaning was given first, it received a weight of "3"; when it was given second, a "2"; when given third, a "1"; and when not given at all, zero weight. The weights for each statement of meaning of a word were averaged across Ss, which yielded an index of relative strength by which the two meanings in the test could be compared.

TABLE II

THE WORDS USED IN EXPERIMENT I, THE MEAN WEIGHTED SCORES FOR THE TWO
MEANINGS, AND PERCENTAGE OF THE JUDGES WHO AGREED ON WHICH
IS STRONGER

	Weighted score[a]	Percentage agreement on which is stronger
Rare		
1. Partially cooked	1.15	76
30. Uncommon	2.64	
Tip		
2. Private information	.77	73
34. Money	1.59	
Diamonds		
3. Red spots on playing cards	.03	97
36. Precious stones	2.82	
Cross		
5. A religious symbol	1.49	89
28. An X mark	.28	
Yard		
6. A grassy place	2.44	69
29. Three feet	1.54	
Board		
7. Meals	.69	89
32. Flat piece of wood	2.62	
Corn		
8. A little lump on someone's foot	.13	100
35. Plants in a field	3.00	
Pen		
21. A fenced enclosure	.72	97
40. Writing implement	2.90	
Palms		
22. Part of a person's hands	2.03	50
39. Plants or trees	2.23	
Crack		
23. Hole or crevice in a wall	2.46	92
45. A sarcastic remark	.49	
Pit		
24. Hole in the ground	2.79	92
44. Hard stone of a fruit	.67	
Toast		
27. A drink in honor of someone	1.85	70
43. Heated and browned bread	2.54	
Bats		
11. Flying animals	1.54	64
33. Wooden sticks	2.28	
Deck		
14. Part of ship	2.92	92
38. A pack of playing cards	.92	

TABLE II (Continued)

		Weighted score[a]	Percentage agreement on which is stronger
Date			
15.	Appointment to take a girl out	1.72	92
37.	A piece of fruit from a palm tree	.51	
Racket			
16.	Noise	1.26	52
25.	Dishonest business	1.03	
Bark			
17.	Sound made by a dog	2.23	53
42.	Outer covering of a tree	2.41	
Bank			
19.	A financial establishment	2.85	87
41.	Side of a river	1.56	
Fall			
20.	A season of the year	1.05	90
47.	To trip and hurt oneself	2.82	

[a] Higher score means stronger meaning.

Table II shows the mean strength of the two meanings represented by the alternatives in the test items, together with the percentage of judges who concurred in the group decision as to which is strongest. As seen there, the judges showed high agreement for some words, and lower for some others.

The prediction was that schizophrenics, more often than normal Ss, would make misinterpretations mediated by the stronger normal meaning response when the weaker meaning response was appropriate in the context of the sentence in which the word appeared. However, one would also expect schizophrenics sometimes to make misinterpretations mediated by a meaning response which according to the group norms is weaker than the correct one. This should occur because, as seen in Table II, there is not complete uniformity among people as to which is the stronger meaning response. However, misinterpretations in this direction should be fewer than those in the predicted direction, and they should occur more often on the words for which the judges show lower agreement as to which is the stronger meaning.

Results. The means of the corrected scores for misinterpretations mediated by the stronger and by the weaker meaning responses are shown in Table III. These data were analyzed using *t* tests. Double-tailed probability values are reported. The schizophrenics exceeded the normal Ss in the wrong interpretations mediated by stronger meaning responses ($z = 5.13$; $p < .001$) and in wrong interpretations mediated by weaker meaning responses ($z = 2.46$; $p < .02$). However, on a score of the dif-

ference between the scores for the two kinds of misinterpretation the schizophrenics again exceeded the normal Ss ($z = 4.22$; $p < .001$), owing to the schizophrenics' differentially greater reliance on the stronger meaning responses.

TABLE III
MEAN NUMBER OF MISINTERPRETATIONS MEDIATED BY STRONGER AND WEAKER MEANING RESPONSES

	Total groups	
	Schizophrenic	Normal
Stronger	3.80	.89
Weaker	1.24	.44
Strong minus Weak	2.56	.45

Superficially, the finding that the schizophrenics exceeded the normal Ss in misinterpretations mediated by weaker meaning responses might appear contradictory to the theory. This finding is nevertheless consistent with the theory if the items which contributed to the difference were those on which the judges were less unanimous as to which is the stronger meaning. The logic is that if a meaning response which is weaker by the group judgmental norms is nevertheless the stronger of the two meaning responses to this word for a minority of the judges, it is also stronger for a minority of the schizophrenics. To test this, the 19 items for which the weaker meaning response mediated the correct answer were divided into two sets, 10 for which the judges had relatively high agreement (89% to 100%) on which was stronger, and 9 for which the judges had relatively low agreement (50% to 87%). (The items in these sets may be identified in Table II). As seen in Table IV, the dif-

TABLE IV
MEAN NUMBER OF MISINTERPRETATIONS MEDIATED BY WEAKER MEANINGS FOR 10 ITEMS WITH HIGH AGREEMENT ON RELATIVE STRENGTH AND 9 ITEMS WITH LOW AGREEMENT

	Schizophrenic	Normal
Low Agreement	1.00	.22
High Agreement	.24	.22

ference between the normal and schizophrenic Ss on misinterpretations mediated by weaker meaning responses was almost completely accounted for by the nine items on which the judges tended to disagree. The difference between groups was significant ($z = 2.79$; $p < .01$) for these nine items. For the remaining 10 items on which the judges showed high agreement on the relative strength of the two meanings, schizophrenic and normal Ss were almost identical on the score of misinterpretation mediated by weaker meaning responses, as predicted by the theory.

One might wonder whether the finding that schizophrenic misinter-pretations are mediated more by stronger meaning responses than by weaker ones might merely reflect a lower intellectual functioning of the schizophrenics. Disturbed chronic schizophrenics show a lowered ability in handling any complex task—even a vocabulary test like the present one. The items were such that when the wrong interpretations were mediated by stronger meaning responses, the correct interpretations were mediated by weaker ones. It may require greater intellectual ability to distinguish the correct answer when it requires mediation by a weaker meaning response than when it requires a stronger one. In order to check on this possibility, it seemed most suitable to use another vocabulary test since the experimental instrument was itself a vocabulary test.

It was possible to obtain Stanford–Binet vocabulary scores for all of the normal Ss and for 21 of the 25 schizophrenics (the remaining four patients being untestable on this instrument). It is most appropriate to test for the relationship in the normal group alone, since the lower vocabulary score of the schizophrenics may reflect inability to take the test in addition to lowered knowledge of words. For normal Ss the score, for the entire test, of the difference between the two kinds of misinter-pretation had only a slight nonsignificant relationship to Stanford–Binet vocabulary score ($r = -.13$) in the direction of a high difference score tending to accompany a low vocabulary score. Further evidence against the importance of vocabulary skill in accounting for the results is that the mean Stanford–Binet vocabulary score for the 21 schizophrenics was, as shown in Table I, only slightly lower than that of the normal Ss, and the difference did not approach significance ($z = .75$; $p = .46$). Nevertheless, in order to make certain that the difference between nor-mal and schizophrenic groups on the difference score was independent of vocabulary skill, an analysis of covariance using both normal and schizophrenic Ss was computed in which the effects of Stanford–Binet vocabulary were partialed out. The schizophrenic and normal groups were again found to differ in the predicted direction on the difference score ($F = 22.59$; $df = 1.45$; $p < .001$).

It seems clear, therefore, that we are justified in concluding not only that schizophrenics' excessive misinterpretations of double-meaning words are more often mediated by the stronger than by the weaker meaning responses, but that in addition this error propensity cannot be attributed to lowered vocabulary skill.

C. Experiment II. Errors of Exclusion from Common Concepts

While schizophrenic thought disorder has many striking features, perhaps the most widely discussed feature of all is the difficulty in

assigning objects to common conceptual classes. The better known discussions of this difficulty include those by Cameron (1939), Goldstein and Scheerer (1941), Hanfmann and Kasanin (1942), McGaughran and Moran (1956), and Vigotsky (1934). It may be that some errors of schizophrenics in the assigning of objects to common conceptual classes arise from reliance on the strongest normal meaning responses either to the name of the conceptual classes or to the names of the objects. The present study investigates only the first of these, errors which reflect excessive reliance on a strong meaning response to a class name, with a neglect of weaker ones.

The conceptual class names used were selected from the many common conceptual classes having more than one meaning. For example, the word *head* may mean either the head of an animate thing, such as of a man or a horse, or the head of an inanimate thing, such as of a nail, a match, or a hammer. However, the animate meaning is the stronger of these two meaning responses to *head*. It was hypothesized that when asked to sort the names of objects into such double-meaning conceptual classes, schizophrenics more than normals would interpret the concepts solely in accordance with the stronger normal meaning response to each class name and would exclude examples mediated by the weaker meaning response.

Method. The S was required to sort a series of cards, each marked with the name of one object on it, into two piles—those which belonged in a certain named conceptual class, and those which did not.

Two boxes with slots in the top were placed in front of the S. A guide card naming the conceptual class was attached to the top of each box.

For Task A, the guide card on one box said "*Things that have a head*" and on the other box, "*Things that do not have a head.*" The sorting cards were of three types, names of living things that have a head (such as *dog* and *horse*), names of inanimate objects that have a head (such as *nail* and *pin*), and irrelevant items.

Task A: The response cards for the concept of "*Things that have a head*" were:

    animate: dog, mouse, horse, pig, cat, sheep, cow, goat.
  inanimate: pin, spear, match, nail, spike, hammer, screw, arrow.
 irrelevant: sheet, blanket, napkin, tablecloth, quilt, towel, bath mat, dish rag.

Task B: The response cards for the concept of "*Things that have legs*" were:

    animate: cow, mouse, rat, horse, goat, cat, lion, pig.
  inanimate: chair, bed, piano stool, table, sofa, foot stool, bench, desk.

irrelevant: pear, banana, peach, apple, orange, grape, pineapple, lemon.

Task C: The response cards for the concept of *"Things that have teeth"* were:

animate: dog, horse, cow, lion, man, alligator.
inanimate: zippers, saw, rake, comb, bear trap, barbers' clippers.
irrelevant: hickory, oak, pine, birch, cedar, maple.

Task D: The response cards for the concept of *"Things that have skin"* were:

animate: goat, elephant, horse, pig, dog, cow, sheep, cat, man, rat.
inanimate: apricot, prune, tomato, potato, apple, cherry, banana, peach, pear, plum.
irrelevant: sled, television set, baseball bat, golf clubs, playing cards, ice skates, radio, checkers, marbles, bowling ball.

Evidence concerning differences in strength of meaning responses to each of these four class names was obtained in the same manner as described in Experiment I. Twenty students in an introductory psychology course were used as judges. Each judge was asked to give aspects of meaning for *head*, *legs*, *teeth*, and *skin*. The judges almost uniformly interpreted all four concepts primarily in terms of the animate examples, although the other meanings were sometimes given. All 20 students gave as their first response a meaning indicating an interpretation in terms of the animate meaning for *head*, for *legs*, and for *teeth*, while 17 of the 20 did so for *skin*.

The prediction was that schizophrenics, more than normals, would tend to respond to the class names on the experimental instrument in terms of the animate meaning only, and therefore, more than normals would exclude names of inanimate examples from the conceptual classes.

However, random placements would be expected to inflate the number of exclusions of both types by the schizophrenics. For example, schizophrenics would be expected more often than normals to sort carelessly, or to have difficulty reading the cards, or to sort according to box position. Such errors should result in excluding both animate and inanimate representatives from the named conceptual class even though such "random" errors do not truly reflect the patients' interpretation of the conceptual class. The number of "irrelevant" cards placed in the named conceptual class was used as a measure of such random error in the same manner as the number of markings of "unidentified error alternatives" was used in Experiment I as a correction for random marking. Corrected scores of animate exclusions and inanimate exclusions were computed

by subtracting the number of the incorrect placements of the irrelevant cards from the number of both the animate and inanimate exclusions.

Results. Between-group comparisons were made using a *t* test. Probability values corresponding to double-tailed tests are reported.

The mean corrected scores of animate and inanimate exclusions are shown in Table V. The results were very much as hypothesized. As seen

TABLE V

MEAN CORRECTED SCORE OF EXCLUSIONS OF ANIMATE AND INANIMATE
REPRESENTATIVES FROM THE CONCEPTUAL CLASSES

	Schizophrenic	Normal
Animate	.16	.19
Inanimate	11.61	3.81

in Table V, the schizophrenics exceeded the normal Ss on the number of exclusions of the inanimate representatives from the named class ($z =$ 3.07; $p < .01$) but were about equal to the normal subjects on exclusions of the animate members. The score of the difference between the number of the two kinds of exclusions distinguished the two groups ($z =$ 4.06; $p < .001$). This indicates that the schizophrenics tended more than the normal Ss to narrow their interpretation of the conceptual classes to that one interpretation mediated by the strongest normal meaning response.

As a check on whether these results might reflect lowered vocabulary ability in schizophrenia, the difference score was correlated with Stanford–Binet vocabulary score for the normal Ss. The correlation was .03, which indicates that no relationship exists and the schizophrenic error pattern was not a reflection of lowered knowledge of vocabulary. Using subgroups matched on vocabulary score, the schizophrenics were again found to exceed the normals on this difference score ($z = 2.82$; $p < .01$).

D. EXPERIMENT III. THE INFLUENCE OF STRONG CONTEXTUAL CUES

Since in the experiments presented above the patients' performance showed some failure of mediation by the weaker meaning responses, one might wonder whether this reflected an absolute loss, in at least some of the patients, of ability to produce responses mediated by some of the weaker meaning responses regardless of the contextual cues. According to the theory, this is not the case. The theory states that the deficit in performance arises from an excessive reliance on the strongest mediating responses regardless of whether the mediating responses are aroused by the word itself or by a contextual cue. Therefore, it follows that if the contextual cues that indicate the appropriateness of the weaker meaning response are extremely strong, and contextual cues for using the strongest

meaning response are absent, the schizophrenics' use of the weaker meaning response will approach that of the normals. If, contrary to the theory, schizophrenia is accompanied by an absolute loss of some weaker meaning responses, then schizophrenics should show a deficit in their use regardless of the context. Experiment III provides a comparison of schizophrenic and normal performance on ability to use weaker meaning responses under these conditions.

Method. The task used was a multiple-choice vocabulary test for double-meaning words. The task required S to select the correct meaning from among four alternatives, of which the first two alternatives included both a correct and an incorrect answer, the third stated "neither of the above," and the fourth, "I don't know." There were two items for each multimeaning word, one for each of two meanings. For example, the word *bear* was presented in two items.

28. The word *bear* may mean:
    A. to carry
    B. to command
    C. neither of the above
    D. I don't know.
64. The word *bear* may mean:
    A. a sharp end
    B. an animal
    C. neither of the above
    D. I don't know.

The meanings corresponding to the stronger and weaker meaning responses and their relative strength were identified by the same procedure as used in Experiment I. This measure indicated that the meaning of *bear* corresponding to the stronger meaning response is that of "animal." If some schizophrenics have an absolute loss of some of the weaker meaning responses, the schizophrenic group should show lower accuracy than normal Ss on items calling for weaker meaning responses to words than on items calling for stronger ones. However, the presentation of the statement of the weaker meaning as one of the alternatives furnishes an extremely strong contextual cue which indicates that the correct response is mediated by that weaker meaning response. The choice of the alternative corresponding to the weaker meaning is actually mediated by a very strong mediating response in this case. Therefore, the theory predicts that schizophrenics would use this strong mediating response to give the correct answer corresponding to the weaker meaning.

The experimental instrument consisted of 130 items of which 74 (the experimental items) were constructed from 37 double- or multi-meaning words, i.e., 2 items per multimeaning word. There were also 46 items constructed for single-meaning words and 10 filler items for

TABLE VI
The Words Used in Experiment III and the Mean Weighted Scores for the Two Meanings

	Weighted score[a]
Barrel	
1. Metal tube on a gun	.36
31. Large container	2.85
Ruler	
2. Measuring stick	2.28
32. A person who governs	1.59
Trunk	
6. Part of an elephant	.56
42. The main stem of a tree	.59
Deal	
7. A bargain	2.56
44. To pass out playing cards	1.62
Log	
9. Piece of wood	2.87
43. The daily record of a ship's voyage	.67
Deck	
70. A pack of playing cards	.92
90. The floor of a ship	2.92
Date	
88. A sweet fruit from a kind of palm tree	.51
126. Appointment with a girl	1.72
Spring	
26. A season of the year	2.15
52. To leap or jump	.69
Run	
27. To move the legs quickly	2.92
53. To be a candidate for election	.05
Bear	
28. To carry	.36
64. An animal	2.72
Board	
38. A flat piece of wood	2.62
63. Meals provided for pay	.69
House	
41. A building in which people live	2.92
74. An assembly for making laws	.05
Note	
46. A very short letter	2.49
73. A musical sign	.38
Jar	
48. A container, as of glass	2.49
81. To shake or rattle	1.15

TABLE VI (*Continued*)

	Weighted score[a]
Seal	
51. A kind of sea animal with flippers	1.10
77. A design stamped on a piece of paper	.28
Plate	
57. A dish	2.85
85. The home base in baseball	.15
Mars	
60. A planet	2.82
83. A god of war	.31
Point	
61. A sharp end	1.85
87. To aim	1.18
Pole	
66. A long slender piece of wood	2.59
98. An end of a magnet	.64
Star	
67. A heavenly body	2.92
97. An outstanding movie actor	.23
Top	
71. The highest point or part	2.82
124. A toy that spins on a point	.28
Bolt	
78. To run away	1.03
123. A strong pin of metal or wood	1.87
Hail	
79. A shout of welcome	1.62
129. Frozen rain	1.82
Count	
82. A European nobleman	1.03
122. To add up	2.69
Sign	
84. To write one's name	1.62
125. A billboard	2.03
Club	
89. A heavy stick	2.08
110. A social group	2.44
Iron	
92. A metal	1.90
108. To press clothing	1.64
Bowl	
93. A hollow rounded dish	2.18
109. To play a game with a ball and pins	1.23
Box	
94. A container	2.87
116. To fight with the fists	.64

TABLE VI (Continued)

Horn	Weighted score[a]
99. A growth on the head of cattle and deer	.36
115. A kind of musical instrument	2.82
Fire	
100. A flame	3.00
114. To discharge from a job	.15
Lock	
101. A portion of hair	.08
117. A means of fastening doors	2.90
Saw	
102. Looked at	.82
119. A tool for cutting	2.46
Fine	
104. A sum of money paid as a punishment	1.82
127. Excellent	2.15
Light	
105. Not heavy	.26
128. To set fire to	.13
Racket	
111. A loud noise	1.26
121. A dishonest way of making money	1.03
Grain	
112. Seed of plants, like wheat	2.62
120. Little lines and markings in wood	.54

[a] Higher score means stronger meaning.

which a correct alternative was not listed among the four alternatives. The latter two kinds of items were included to prevent Ss from discovering the nature of the task and to reduce the set for never marking the last two alternatives. Table VI lists the 37 words used in the experimental items, together with the meanings offered in the alternatives, and the mean strength of the two meanings.

Results. For this task, unlike those of Experiments I and II and the earlier study reported in Section I, B, the score measuring differential use of stronger and weaker meanings was related to vocabulary skill in normal Ss. The correlation of Stanford–Binet vocabulary score with the difference between the number of errors on the items requiring use of weaker and stronger meanings for the normal Ss was −.48 (p < .05), which means that the brighter Ss showed less difference in their error scores on the two kinds of items. As seen in Table I, the schizophrenics scored slightly lower than the normal Ss on Stanford–Binet vocabulary. It seemed inappropriate, therefore, to use the entire normal and schizophrenic groups in testing for the experimental effect since the predicted

difference might be found simply as a reflection of the lowered vocabulary skill of the patients. For this reason, the experimental effects were evaluated using subgroups of 20 normal and 20 schizophrenic Ss closely matched on Stanford–Binet vocabulary score, as shown in Table I.

Table VII shows the median number of errors for these matched subgroups on the items calling for answers mediated by stronger and weaker meaning responses. As seen there, the scores were quite similar for the two groups, although qualitatively the pattern of scores was slightly in the direction opposite to that which would result from a schizophrenic loss of weaker meanings. The difference between the groups on the number of weaker meanings failed minus the number of stronger meanings failed did not approach significance ($z = .18$; $p = .95$) as determined by a double-tailed Mann–Whitney test.

TABLE VII

MEDIAN ERROR SCORES ON ITEMS FOR WHICH STRONGER AND WEAKER MEANING RESPONSES YIELD CORRECT ANSWERS

	Schizophrenic	Normal
Stronger	.79	.70
Weaker	2.07	2.21

These results are in striking contrast to those obtained in Experiments I and II. It appeared possible that this difference in results could be a consequence of differences in the stimulus words instead of the hypothesized differences in strength of contextual cues. For example, if the weaker meanings of the words in Experiment III were not quite as weak as those in Experiment I, the schizophrenics might show less difference between the two kinds of items in Experiment III for that reason alone. This possibility can be evaluated by examining the mean strength of the stronger and weaker meaning responses as listed in Tables II and VI. In Experiment I, the mean strengths of the stronger and weaker meanings were 2.41 and 1.01, respectively; in Experiment III, 2.33 and .77, respectively. Thus, in Experiment III, the weaker meanings are qualitatively weaker (.77 as compared with 1.01) than in Experiment I and the difference in the stronger and weaker meanings is greater for Experiment III (1.56 compared to 1.40). These differences in relative strength are slightly in the direction opposite to that which might account for the differences in the results. Therefore, it seems most likely that the differences in the results obtained with the two tasks are attributable to the differences in the strength of the contextual cues.

These findings are inconsistent with the interpretation that schizophrenics have lost access to weaker meaning responses. However, the findings do support the contention of the present theory that schizophrenics' overt responses, more than those of normal persons, are medi-

ated by the strongest mediating responses, regardless of whether they are aroused by the stimulus words or by contextual cues.

III. Discussion

Although the theory emerged from other research (Chapman and Chapman, 1964), the primary data of these three studies were gathered prior to the formulation of the theory, and only the measures of meaning responses were gathered with the theory in mind. Therefore, it was not possible to design the experiments so as to maximize the hypothesized effects. It would have been easier to support the theory using only those double-meaning words for which there is a sharp differentiation between the strengths of strong and weak meaning responses. Also, ideally, the same stimulus words should have been used in Experiments I and III. However, the support of the theory by the data is perhaps more convincing since they were not gathered in a manner designed to support it.

The theory generates predictions, enumerated below, concerning the conditions under which a wide range of well-known features of schizophrenic disorder of thought will occur and will not occur. Insofar as these predictions should be confirmed in future research, we will be able to conclude that the theory subsumes these various features of thought disorder under a single explanatory principle.

In the studies reported in the present paper it was found that the theory predicts the ways in which schizophrenics misinterpret double-meaning words presented in context. This kind of misinterpretation is very similar to the confusion of usages of words and phrases which have both a literal and figurative (or metaphoric) usage. This kind of confusion has recently received extensive attention from psychopathologists. Bateson, Jackson, Haley, and Weakland (1956) view the confusion of metaphoric and literal usages as an important feature of schizophrenic thinking. The present theory would predict that schizophrenics will make such misinterpretations under the same conditions as they misinterpret other double-meaning words, i.e., they will choose the interpretation mediated by the strongest normal mediating response regardless of whether it is figurative or literal.

Several investigators have been interested in the observation that schizophrenics seem to have a special difficulty understanding humor. A great many jokes are dependent on double meanings of words, and seeing the point of the joke requires mediation of overt response by both strong and weak meanings of the word. The theory, therefore, yields clear predictions concerning what kinds of jokes schizophrenics should have special difficulty in understanding.

The theory also generates predictions which may resolve a seeming contradiction between clinical observations and experimental evidence concerning errors in syllogistic reasoning. Both von Domarus (1944) and Arieti (1955) have reported that an especially striking feature of schizophrenic disorder of thought is a tendency to conclude, in syllogistic reasoning, that two objects are identical if they share a common quality. However, Gottesman and Chapman (1960) found no differences between schizophrenics and normals on a measure of such errors. The present theory would predict that the von Domarus error is stronger in schizophrenics than in normal subjects only when the quality shared by the two objects corresponds to a strong mediating response to each of them. This was not the case for the syllogisms used in the Gottesman and Chapman study. If further research indicates that the von Domarus phenomena occur only when the shared quality corresponds to a high-strength meaning response, it will appear that such phenomena are better accounted for by the present theory.

The theory also predicts that schizophrenics should show abnormally great semantic generalization or abnormally great transfer in a verbal learning task across some kinds of words and not others. The heightened semantic generalization or transfer should occur across pairs of words which share meaning responses of highest strength but which also have other meaning responses which are not shared. For pairs of words which do not fit this specification, schizophrenics should not show excessive semantic generalization or transfer effects.

The theory also accounts for some of the behaviors, such as schizophrenics show on conceptual sorting tests, which are called "concrete" thinking by various writers, e.g., Goldstein and Scheerer (1941). For example, on one task of the object-sorting test the subject is instructed to sort objects together which "belong together," and on another task he is presented objects already grouped together and is asked, "Why do these objects belong together?" The theory predicts that schizophrenics' sorting on the first task and explanations of belongingness in the second will be mediated to an abnormal degree by the strongest meaning responses to the objects. This means that the schizophrenics' grouping of objects and the principle of sorting that they offer in response to a question concerning belongingness should show mediation by normal meaning response of highest strength regardless of whether the response would be labeled "abstract" or "concrete." For example, the theory predicts that schizophrenics will show an excessive use of sensory qualities only for sets of objects to which the strongest normal mediating response is that sensory quality.

The data of Experiment II showed that schizophrenics' interpreta-

tion of the conceptual class name in accordance with its strongest meaning response may produce errors of overexclusion of objects from the conceptual class. The theory may also account for errors which appear superficially to be an opposite kind of error, that is, errors of overinclusion, and it again predicts the kinds of items on which such errors will occur. Schizophrenics should exceed normal persons in the inappropriate inclusion of names of objects in a conceptual class when the strongest normal meaning response is shared by the incorrect objects and objects that properly belong in the conceptual class.

Another advantage of the present theory is that it might be used to make predictions concerning individual differences among schizophrenics as to the items on which they make various errors. This could be done by obtaining meaning responses from the patients themselves for the words or objects which are used on the tests.

In comparison to other theories of disordered thought in schizophrenia, the present theory is both broad in the range of its predictions and precise in the nature of its predictions. It is precise in that it predicts a certain kind of error for each situation, and predicts the kind of stimuli with which that kind of error will appear and will not appear. A major virtue of the theory is that its predictions are readily amenable to experimental test.

Even if all of these predictions should be confirmed, the present writers would not wish to claim that the theory accounts for *every* deviant verbal production of schizophrenics. It may be that other principles, as well as the present one, are necessary to produce an adequate descriptive system. Also, while the theory and data presented here have been in terms of meaning responses held in common by both schizophrenics and normals, it is possible that in addition schizophrenia produces some increase of deviant meaning responses. Also, this theory does not go beyond description to attempt an explanation of the etiology of the disorder. Nevertheless, the usual first step in understanding a disorder is to describe it, and the chief attraction of the present theory is the promise that it holds for reducing the number of principles necessary for that description.

References

Arieti, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.

Bateson, G., Jackson, D. D., Haley, J., and Weakland, J. H. Toward a theory of schizophrenia. *Behav. Sci.*, 1956, 1, 251-264.

Bleuler, E. *Dementia praecox or the group of schizophrenias.* Translated by Joseph Zinkin. New York: International Univer. Press, 1950. Originally published 1911.

Bousfield, W. A., Whitmarsh, G. A., and Danick, J. J. Partial response identities in verbal generalization. *Psychol. Rep.*, 1958, 4, 703-713.

Cameron, N. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939, 85, 1012-1035.

Campbell, D. T. Systematic error on the part of human links in communication systems. *Information and Control*, 1958, 1, 334-369.

Chapman, L. J. Intrusion of associative responses into schizophrenic conceptual performance. *J. abnorm. soc. Psychol.*, 1958, 56, 374-379.

Chapman, L. J. Confusion of figurative and literal usages of words by schizophrenics and brain-damaged patients. *J. abnorm. soc. Psychol.*, 1960, 60, 412-416.

Chapman, L. J. A re-interpretation of some pathological disturbances in conceptual breadth. *J. abnorm. soc. Psychol.*, 1961, 62, 514-519.

Chapman, L. J. The problem of selecting drug-free schizophrenics for research. *J. consult. Psychol.*, 1963, 27, 540-542.

Chapman, L. J., and Chapman, Jean P. Atmosphere effect re-examined. *J. exp. Psychol.*, 1959, 58, 220-226.

Chapman, L. J., and Chapman, Jean P. The interpretation of words in schizophrenia. *J. abnorm. soc. Psychol.*, in press, scheduled for publication in 1964.

Chapman, L. J., and Knowles, R. R. The effects of phenothiazine on disordered thought in schizophrenia. *J. consult. Psychol.*, 1964, 28, 165-169.

Chapman, L. J., and Taylor, Janet A. Breadth of deviate concepts used by schizophrenics. *J. abnorm. soc. Psychol.*, 1957, 54, 118-123.

Deese, J. On the structure of associative meaning. *Psychol. Rev.*, 1962, 69, 161-175.

Goldstein, K., and Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2, Whole No. 239.

Gottesman, L., and Chapman, L. J. Syllogistic reasoning errors in schizophrenia. *J. consult. Psychol.*, 1960, 24, 250-255.

Hanfmann, E., and Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, No. 67.

McGaughran, L. S., and Moran, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and non-psychiatric groups. *J. abnorm. soc. Psychol.*, 1956, 52, 43-50.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

Underwood, B. J. An orientation for research on thinking. *Psych. Rev.*, 1952, 59, 209-220.

Vigotsky, I. S. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, 1934, 31, 1063-1077.

von Domarus, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California, 1944, pp. 104-113.

THE EFFECT OF THE EXPERIMENTER ON THE RESULTS OF PSYCHOLOGICAL RESEARCH[1]

Robert Rosenthal

HARVARD UNIVERSITY, CAMBRIDGE, MASSACHUSETTS

The purpose of this paper is to invite consideration of the experimenter himself as a source of variance in the results of psychological experiments. As different as experiments may be from one another with respect to purposes, procedures, and subject sampling, they all have in common that someone must collect the data—the experimenter or his surrogate. To the extent that the experimenter himself may be a significant determinant of the results of his research, it may become necessary to re-evaluate carefully the results of experiments completed and the design of experiments proposed.

We may distinguish two broad types of effects which experimenters may have upon their results: (1) those which operate without directly affecting the subjects of the experiment and (2) those which lead to subjects changing the nature of their responses.

I. Effects Not Altering Subjects' Responses

A. OBSERVER EFFECTS

In one way or another the experimenter must observe the subjects' responses. By observer "effects" or "error" we mean over- or underestimation of some criterion value. When two observers disagree in an observation, each may be said to err with respect to the other. Both may be said to err with respect to some third observation which may, for various reasons, be a more or less usefully employed criterion. Given a population of observations, we may choose to define their mean value as the "true" value and regard all observations not falling at the mean as being more or less in error as a direct function of their distance from the mean.

Observer errors or effects may be distinguished from observer "bias" by the fact that observer errors are randomly distributed around a "true" or "criterion" value. Biased observations tend to be consistently too high or too low and may bear some relation to some characteristics of the observer (Roe, 1961), the observation situation (Pearson, 1902), or both.

In considering the act or sequence of acts constituting the observation in the scientific enterprise, we may distinguish conceptually among locations of error or bias. The error of "apprehending" occurs when some sort of misrecording occurs between the event observed and the observer of the event. We may include here such diverse sources of apprehending error as differing angles of observation (George, 1938), imperfections in the sensory apparatus, relay systems, cortical projection areas, and the like. The error of recording may be conceptually distinguished from the apprehending error. We assume an errorless act of apprehending but a transcription of the event to paper, to the ear of another observer, or to another instrument which differs from the event as correctly apprehended. In actual practice, of course, when an event or observation is recorded in error with respect to some criterion, we cannot locate the error as having occurred in apprehending, in transcribing, or in both processes. There is no way we can isolate an apprehending error unconfounded with a recording error. Computational errors are more clearly distinguishable from the foregoing errors since they involve the incorrect manipulation of recorded events. Incorrectness is usually defined here by the formal rules of arithmetic operations.

The magnitude and extent of observer effects are well known in the

history of science (Rosenthal, 1963a). In our own discipline, Boring (1950) has made us particularly aware of observer variance by his discussion of the personal equation of the astronomers. Most contemporary discussions of observer effects, however, seem to have become more perfunctory. Reference is made to modern instrumentation which may serve to eliminate observer error. That observer effect may be reduced by mechanical means seems reasonable enough. That instrumentation may not eliminate observer effect must also be considered. If the instrument be a dial, it must be read by a human observer. If the instrument be a computer, the printout must likewise be read by an observer. Observer effect or variability in the reading of scales was well noted by Yule (1927). A general error tendency which he found was the inclination to read scales to quarters of intervals rather than to tenths. Empirical analysis of his own observer effect revealed to Yule his tendency to avoid the number 7 as a final digit and to favor the numbers 8, 9, 0, and 2. Yule, incidentally, was relatively optimistic that proper training could eliminate observer error. The equally reknowned methodologist R. A. Fisher (1936) did not, however, seem to share this optimism.

There have been few investigations of observer effects in the psychological experiment. A notable exception was the work of Sheffield, Kaufman, and Rhine (1952). In an experiment in psychokinesis, they filmed the actual fall of the dice which subjects were trying to influence. They found subjects believing in the phenomenon to make tallying errors in favor of the hypothesis. Subjects who disbelieved made the opposite type of tallying error.

Both recording as well as computational errors of experimenters were systematically and explicitly studied for a person perception task experiment (Rosenthal, Friedman, Johnson, Fode, Schill, White, and Vikan-Kline, in press). In that study, each subject wrote on a pad his rating of the degree of success or failure of persons pictured in photographs. The 30 experimenters of this study transcribed these ratings to a master data sheet during the course of their interaction with their subjects. A comparison of experimenters' transcriptions with subjects' recordings revealed only 20 errors out of 3000 data transcriptions. All errors were made by 12 of the 30 experimenters. Errors were not randomly distributed but tended to fall in the direction of the experimenters' hypotheses.

The computation task for the experimenter in the study under discussion was simply to sum the 20 ratings given him by each of his 5 subjects. Most experimenters erred computationally and nonrandomly so. Not only were more biased experimenters more likely to make computational errors in their favor; they also tended to make larger computa-

tional errors. In this same experiment, all subjects rated their experimenters on the variable of "honesty" during the conduct of the experiment. The mean rating was quite high on honesty and the range of ratings quite small. Experimenters could not actually have been "dishonest" even if they had been so inclined. During the experiment, the co-investigators had all experimenters under surveillance (a fact apparent to all experimenters). In spite of the restricted range of the ratings of the experimenters' honesty made by the subjects, these ratings predicted far better than chance whether the experimenter would subsequently favor himself more in the making of a computational error.

In the study under discussion, we found a significant correlation (.48, $p = .01$) between the occurrence of recording errors and the occurrence of computational errors. Thus, experimenters who erred in data transcription tended to err in data processing. However, those experimenters who erred in the direction of their hypothesis in their recording errors were not any more likely to err in the same direction in their computational errors. Numerical errors seemed then to be a consistent characteristic while directionality of error vis-a-vis hypothesis did not.

We should note here that the over-all effects of both recording and computational errors on the grand means of the different treatment conditions of the experiment were negligible. An occasional experimenter did have some real effect on the data he obtained; an effect which could be serious if an entire experiment depended on an experimenter who was prone to err numerically. Only rarely in the psychological literature, however, is attention called to a real or alleged numerical error (Hanley and Rokeach, 1956; Wolins, 1962).

B. Intentional Effects

Intentional error production on the part of the experimenter is probably as rare an event in the psychological experiment as it is in the sciences generally (Wilson, 1952; Shapiro, 1959; Turner, 1961). Nevertheless, any serious attempt at understanding the social psychology of psychological research must consider the occurrence, nature, and control of this type of experimenter effect.

In survey research the "cheater problem" is of sufficient importance to have occasioned a panel discussion of the problem in the *International Journal of Attitude and Opinion Research* (1947). Such workers as Blankenship, Connelly, Reed, Platten, and Trescott seem to agree that, though statistically infrequent, the cheating interviewer can affect the results of survey research, especially if the dishonest interviewer is responsible for a large segment of the data collected. Other workers

calling attention to the cheating problem in survey research include Crespi (1945–1946), Mahalanobis (1946), and Cahalan, Tamulonis, and Verner (1947). One way in which the problem of intentional error in survey research differs from the problem in the sciences generally is in the status, career aspirations, and identifications of the data collectors. While most data in the sciences are gathered by professional scientists and their professional assistants, the interviewers in survey research are often part-time nonprofessionals with less education and less aspiration, if any, to identification with a scientific career role.

A similar situation exists in the laboratory exercises performed as part of undergraduate courses in experimental psychology. In such laboratory exercises the fabrication of data is commonplace and well known to instructors. Positive sanction is sometimes involved for the fabrication of "real-appearing" data. The same sort of fabrication occurs in undergraduate physics laboratory courses. Students in these undergraduate laboratory courses, like interviewers in survey research, do not generally have the same commitment to "science," however, that professional scientists or research assistants, scientists in the making, have. In two experiments carried out by undergraduate laboratory students, attempts were made to check the incidence of dishonesty. In both situations, the experiments performed were structured not as just "laboratory exercises" but rather as sources of "real data" to be used in the publications of "real scientists." In one of these experiments involving the learning of rats in Skinner boxes, several instances of data fabrication came to light (Rosenthal and Lawson, 1963). In the other experiment, through the use of participant observers, a somewhat more precise estimate of intentional error production was possible (Rosenthal and Fode, 1963a). In that study of discrimination learning in rats, no instances of actual data fabrication were noted, though it is remotely possible that they might have occurred. Rather, error production involved the use of prodding of animals on a number of occasions. These "errors" were randomly enough distributed over treatment conditions that the obtained means, and more importantly, the differences between means, were probably not seriously affected by their occurrence.

Somewhat closer to the situation of the professional, science-oriented, data collector is the situation described by Azrin, Holz, Ulrich, and Goldiamond (1961). They replicated a verbal conditioning study originally conducted by Verplanck (1955) and obtained results similar to his. In both studies, experimenters had been able to control the content of their subjects' verbal productions by verbal reinforcement. It developed, however, that the formal procedure for experimenters was such that, if followed, the experimental effects could not be demonstrated! A number

of the experimenters, it was learned, had resorted to the fabrication of the data they believed the investigators to want.

The Control of Intentional Errors

In discussing the control of intentional errors committed by survey interviewers, Crespi (1945–1946) stated that these errors were more a problem of morale than of morals. After discussing various ways of assessing the data for the likelihood of cheating, he returns to the theme that most of all, we need to know why data collectors cheat. Removal of the conditions motivating intentional error production seems to be the most valuable control of their occurrence. No method of *post hoc* analysis (*International Journal of Attitude and Opinion Research*, 1947) of whether cheating occurred can ever be a satisfactory substitute for eliminating the need to cheat.

Perhaps one long-range improvement in data-collector attitudes would be the revision of the current practice in undergraduate laboratory courses of telling students what data they will obtain. Better put a high reward on data honestly, if unskillfully, obtained than to reward the "correct" data, skillfully but dishonestly manufactured. There seems to be something unresearchman-like in telling students what they will find at the end of a journey whose purpose is really to confront a surprise.

II. Effects Altering Subjects' Responses

A. EXPERIMENTER ATTRIBUTES

There are now dozens of studies which show quite clearly that subjects' responses may be in part determined by the characteristics of the experimenter who is interacting with the subjects (Rosenthal, 1963b). The experimenter's sex, race, religion, status, likeability, and warmth all have been shown to affect subjects' responses. Prior acquaintance with subjects, adjustment, hostility, anxiety (Winkel and Sarason, in press), acquiescence, authoritarianism, and intelligence are other experimenter variables found to make a difference (Rosenthal, 1963b). A concrete example of a recent experiment investigating the effect of experimenter attributes follows.

The study employed 40 experimenters (almost all male) and 230 subjects, about half of them male and half of them female (Rosenthal, Persinger, Vikan-Kline, and Mulry, 1963e). Each experimenter requested about six subjects to rate the apparent success or failure of twenty persons pictured in photographs. Thirty-one of the forty experimenters took both the Marlowe-Crowne Social Desirability Scale and the Taylor Manifest Anxiety Scale before running their subjects. A correlation of .48 ($p = .02$)

was obtained between the experimenters' anxiety scores and the degree of success their subjects saw in the photos of faces. For these same experimenters, the correlation between their need for social approval as measured by the Marlowe-Crowne Scale and the ratings of success of the photos obtained from their subjects was —.32 ($p = .10$). Thus, subjects rated persons as more successful when in the presence of experimenters who were more anxious and had a lower need for social approval. (For this sample, the correlation between experimenters' need for social approval and their anxiety was —.14.)

The results reported seemed clear enough and we were quite un-prepared for the results of a subsequent experiment (Rosenthal, Kohn, Marks, and Carota, 1963a). With 26 male experimenters running a total of 115 female Ss in the same photo-rating task we obtained diametrically opposed results. This time *less* anxious experimenters obtained ratings of greater success from their subjects ($\chi^2 = 7.55, p < .01$). Furthermore, experimenters *higher* in need for social approval obtained more ratings of success ($\chi^2 = 3.85, p = .05$). Apparently then, personality variables may interact with replications in partially determining the subjects' responses. The most obvious difference between the two experiments reported was in their locale. The first study had been conducted at a medium-sized state university in the midwest. The second had been conducted at an east coast private university. Exactly how these differences might have served to lead to opposite results remains a question for further study. It does seem, however, that where an experiment is conducted can make a significant difference in its outcome—even when the principal investigator is held constant.

B. EXPERIMENTER MODELING EFFECTS

Modeling effects are a special case of the effects of experimenter attributes on the data obtained from subjects. The specific attribute involved here is the quality of the experimenter's own performance at the task he subsequently assigns his subjects. The extent to which a subject's task performance is predictable from his experimenter's own performance of the same task is the extent of the modeling effect.

While studies of modeling effects have been conducted for some time in the field and in the clinic, analogous studies in the laboratory have been undertaken only very recently. This may be attributed to a general belief that the greater "rough-and-tumble" of the field and clinic might naturally lead to increased modeling and related effects. Conversely, the greater degree of "control" and "precision" of the laboratory may have led to the view that these extraneous effects are less likely to occur.

The assumptions underlying the development of a program of

research on laboratory experimenters' effects were that (1) there is no discontinuity of control, precision, or operating processes between field studies, clinical studies, and laboratory studies that are a simple function of the site of the research, and (2) that there is no discontinuity between the processes governing the interpersonal experimenter-subject relationship in any sort of social research and the processes governing any other sort of interpersonal relationship.

A total of eight experiments have been conducted which were designed to assess the existence and magnitude of experimenter modeling effects. All of these studies employed the same task. Subjects were asked to rate a series of 10 or 20 photographs on a rating scale of how successful or unsuccessful the persons pictured appeared to be. In all eight studies, experimenters themselves rated the photos before running their subjects. This was accomplished as part of the training procedure, it being most convenient to train experimenters by having them assume the role of subject while the authors acted in the role of experimenter. For each study, modeling effects were defined by the correlation between the mean rating of the photos by the different experimenters themselves and the mean photo rating obtained by each experimenter from all his subjects. The number of experimenters (and therefore the N per correlation coefficient) per study ranged from 10 to 26. The number of subjects per study ranged from 55 to 206. The number of subjects per experimenter ranged from 4 to 20, the mean falling between 5 and 6. In all, 145 experimenters and over 800 subjects were included (Rosenthal, 1963c).

All experimenters were either graduate students or advanced undergraduate students in psychology or guidance. All subjects were drawn from elementary college courses, primarily in psychology but also from education, social sciences, and the humanities. All of the experiments were designed to test at least one hypothesis about experimenter effects other than modeling effects (as, for example, the effects of experimenters' expectancy) on the data obtained by them. All studies, then, had at least two treatment conditions the effects of which would have to be partially transcended by modeling effects.

Table I shows the correlation obtained in each of the eight studies between data produced by experimenters and data later obtained by them from their subjects. In addition to the correlations expressing modeling effects, the number of experimenters on which the correlation is based is shown for each experiment. We see a remarkable inconsistency of obtained correlations, the range of rhos being from $-.49$ to $+.65$. Taken individually, only the correlation of .65 ($p = .001$) differed significantly from zero at even the .10 level, two-tailed. (This particular correlation was the only one based on data not available for closer study by the writer.)

However, the hypothesis that this array of correlations could constitute a set of chance fluctuations from some "true" value of rho had to be rejected ($\chi^2 = 20.3$, $df = 7$, $p = .005$). The same analysis omitting data from experiment 2 gave a χ^2 value of 9.80, $df = 6$, $p = .15$. It seems more likely than not that in different experiments utilizing a person perception task there will be significantly different magnitudes of modeling effects, which for any single experiment might often be regarded as a chance fluctuation from a correlation of zero.

TABLE I

MODELING EFFECTS BY SERIAL ORDER OF EXPERIMENT

Experiment	N	Rho
1	10	.52
2	24	.65
3	12	.18
4	18	.31
5	18	—.07
6	26	—.32
7	12	—.49
8	25	.14

Table I shows the obtained correlations in the temporal order in which the experiments were performed. There was a tendency for later studies to show significantly more negative correlations than the earlier studies. The correlation (rho) between the algebraic magnitude of correlation and order of study was .81, one-tailed, $p = .03$. (Omitting experiment 2, rho was .75, two-tailed, $p = .05$.) It therefore appears that in later studies the modeling effect significantly decreased algebraically. This finding is difficult to explain. From the earlier to the later studies there seemed to be no systematic change in the character of the experimenters, subjects, or treatment conditions. Almost a guess is the suggestion that in later studies experimenters were somewhat more likely to suspect that they themselves were the objects of study and that this recognition may have tended to reverse the direction of the effect. Experimenters, perhaps, were on their guard to avoid any possible sort of effect and modeling effects may have been so assiduously guarded against as to reverse their effects. Evidence for this "reversal of experimenter effects" will be summarized later.

C. EXPERIMENTER EXPECTANCY EFFECTS

All experimenters have some orientation toward the results of their research. Rarely is this orientation one of dispassionate disinterest. Variables are not chosen for inclusion in research by using tables of random numbers. They are, rather, chosen because the experimenter

has certain expectations about the relationship or lack of relationship between the selected variables and certain other variables. A superficial exception to this might be seen in so-called heuristic hunts for relationships, which are perhaps more common to the behavioral sciences. Even here, however, the inclusion of variables is not on a random basis, and certain relationships appear more likely to be found than others.

We now raise the question of whether an experimenter's expectations can affect the data actually obtained in his research. We are not so much concerned here with the problem of choice of experimental design and the fact that certain designs may unintentionally be more or less favorable to obtaining expected or unexpected data. Neither are we concerned with the problem of statistical tests of hypotheses and the fact that uniquely most powerful statistics may unintentionally be employed when the expectation is to be able to reject the null hypothesis, while less powerful statistics may be employed when the expectation is to be unable to reject the null hypothesis. These are interesting questions but will not be considered here. Our usage of "results" or "outcome" will be restricted to the raw data obtained by experimenters from their subjects. What we are talking about then is something very much like Merton's (1948) concept of "self-fulfilling prophecy." One prophesies an event (i.e., an experimental result) and the expectation of it then changes the prophet's (experimenter's) behavior in such a way as to make the predicted event more likely.

Evidence for the operation of these expectancy effects has been drawn from everyday life, from the field of education, from clinical practice, and from the field of survey research (Rosenthal, in press a). Until recently, however, there has been little experimental evidence that self-fulfilling prophesies may operate even in a carefully controlled laboratory research setting. Perhaps the earliest laboratory experiment which involved the systematic manipulation of the experimenters' expectancies was that by Stanton and Baker (1942). Twelve nonsense geometric figures were presented to a group of 200 undergraduate subjects. After several days, retention of these figures was measured by five experienced workers. The experimenters were supplied with a key of "correct" responses, some of which were incorrect. Experimenters were explicitly warned to guard against any bias associated with their having the keys before them and therefore influencing their subjects to guess correctly. Results showed that the experimenter obtained outcomes in accordance with his expectations. When the item on the key was correct, the subject's response was more likely to be correct than when the key was incorrect. In a careful replication of this study Lindzey (1951) emphasized to his experimenters the importance of keeping the keys out

of the subjects' view. This study failed to confirm the Stanton and Baker findings. Another replication by Friedman (1942) also failed to obtain the significance levels obtained in the original. Still, significant results of this sort, even occurring only in one out of three experiments, cannot be dismissed lightly. Stanton (1942) himself presented further evidence which strengthened his conclusions. He employed a set of nonsense materials, 10 of which had been presented to subjects, and 10 of which had not been. Experimenters were divided into three groups. One group was correctly informed as to which 10 materials had been exposed, another group was incorrectly informed, while the third group was told nothing. The results of this study also indicated that the materials which the experimenters expected to be more often chosen were, in fact, more often chosen.

But the most fascinating and most instructive case of experimenter expectancy effects is no doubt that of Clever Hans (Pfungst, 1911). Hans, it will be remembered, was the horse of Mr. von Osten, a German mathematics teacher. By means of tapping his foot, Hans was able to add, subtract, multiply, and divide both integers and fractions. Hans could spell, read, and solve problems of musical harmony. To be sure, there were other clever animals at the time. There was "Rosa" the mare of Berlin, who performed similar feats in vaudeville, and there was the dog of Utrecht, and the reading pig of Virginia. All these other clever animals were highly trained performers who were, of course, intentionally cued by their trainers.

Mr. von Osten, however, did not profit from his animal's talent, nor did it seem at all likely that he was attempting to perpetrate a fraud. He swore he did not cue the animal and he permitted other people to question and test the horse without his own presence. Pfungst and his colleague, Stumpf, undertook a program of systematic research to discover the secret of Hans' talents. Among the first discoveries made were that if the horse could not see the questioner, Hans was not clever at all. Similarly, if the questioner did not himself know the answer to the question, Hans could not answer it either. Still, Hans was able to answer Pfungst's questions as long as the investigator was present and visible. Pfungst reasoned that the questioner might in some way be signalling to Hans when to begin and when to stop tapping his hoof. A forward inclination of the head of the questioner would start Hans tapping, Pfungst observed. He tried then to incline his head forward without asking a question and discovered that this was sufficient to start Hans' tapping. As the experimenter straightened up, Hans would stop tapping. Pfungst then tried to get Hans to stop tapping by using very slight upward motions of the head. He found that even the raising of the eye-

brows was sufficient to stop the tapping. Even the dilation of the questioner's nostrils was a cue for Hans to stop tapping.

When a questioner bent forward more, the horse would tap faster. This added to the reputation of Hans as brilliant. That is, when a large number of taps was the correct response, Hans would tap very, very rapidly until he approached the region of correctness, and then he began to slow down. It was found that questioners typically bent far forward when the answer was a long one, gradually straightening up as Hans got closer to the correct number. For some experiments, Pfungst discovered that auditory cues functioned additively with visual cues. When the experimenter was silent, Hans was able to respond correctly 31% of the time in picking one of many placards with different words written on them, or cloths of different colors. When auditory cues were added, Hans responded correctly 56% of the time.

Pfungst then played the part of Hans, tapping out responses to questions with his hand. Of twenty-five questioners, twenty-three unwittingly cued Pfungst as to when to stop tapping to give a correct response. None of the questioners (males and females of all ages and occupations) knew the intent of the experiment. When errors occurred, they were usually only a single tap from being correct. The subjects of this study, including one experienced psychologist, were unable to discover that they were emitting cues.

Pfungst summarized the difficulties in uncovering the nature of Clever Hans' talents by speaking of "looking for, in the horse, what should have been sought in man." We shall now report the results of a continuing program of research on the effects of experimenters' expectancies on the data they obtain from their subjects. We have then, in a sense, paraphrased and accepted Pfungst's advice. We seek to find in the experimenter a portion of what has most often been sought in the subject.

III. Demonstration of Expectancy Effects: Human Subjects

The basic paradigm for the study of experimenter expectancy effects has been to create two or more groups of experimenters with different expectancies or hypotheses about the data they would obtain from their subjects. Only the very first experiments carried out will be reported in some detail (Rosenthal and Fode, 1963b). Subsequent experiments will simply be summarized.

Method

Fifty-seven photographs of faces ranging in size from 2×3 cm to 5×6 cm were cut from a weekly news magazine and mounted on 3×5 inch white cards. These were presented individually to 70 male

and 34 female students, enrolled in an introductory psychology class at the University of North Dakota. Ss were instructed to rate each photo on a rating scale of success or failure. The scale ran from —10, extreme failure, to +10, extreme success, with intermediate labeled points. Each S was seen individually by one of the Es (RR) who read to each the instructions given below.

Instructions to Subjects. "I am going to read you some instructions. I am not permitted to say anything which is not in the instructions nor can I answer any questions about this experiment. OK?

"We are in the process of developing a test of empathy. This test is designed to show how well a person is able to put himself into someone else's place. I will show you a series of photographs. For each one I want you to judge whether the person pictured has been experiencing success or failure. To help you make more exact judgments you are to use this rating scale. As you can see the scale runs from —10 to +10. A rating of —10 means that you judge the person to have experienced extreme fail-ure. A rating of +10 means that you judge the person to have experienced extreme success. A rating of —1 means that you judge the person to have experienced mild failure while a rating of +1 means that you judge the person to have experienced mild success. You are to rate each photo as accurately as you can. Just tell me the rating you assign to each photo. All ready? Here is the first photo. (No further explanation may be given although all or part of the instructions may be repeated.)"

From the original 57 photos we selected 10 for each sex which met the following requirements: (a) their mean rating was close to zero (between —1 and +1), (b) their distribution of ratings was not signifi-cantly skewed, (c) when the mean ratings for each of the ten cards was summed, they summed to zero exactly. Thus we obtained a set of "neutral" stimulus-value photos which were rated on the average as neither successes nor failures.

Experimenters. Ten of the eleven students in a class in under-graduate experimental psychology served in this capacity. All were psychology majors and three of them were first-year graduate students in psychology. All but two of the Es were males.

Subjects. Ss were 206 students enrolled in a course in introductory psychology (92 males and 114 females). Because Ss were given class credit for participating in the experiment, most of the class volunteered, thus reducing the selective effect of using "volunteer" Ss (Rosenthal, in press b). Each E ran from 18 to 24 subjects.

Procedure. The E's task was structured as a laboratory exercise to see whether they could replicate "well-established" experimental findings as "students in physics labs are expected to do." Es were told to discuss their

project with no one and to say nothing to their *S*s other than what was in the Instructions to Subjects. All *E*s were paid a dollar an hour except that if they did a "good job" they would be paid double; $2.00 an hour. All 10 *E*s received identical instructions except that five *E*s were told that their *S*s would average a +5 rating on the ten neutral photos. The other *E*s were told that their *S*s would average a —5 rating. Thus the only difference between the two groups of *E*s was that one group had a plus mark written in the front of the "5" while the other group had a minus mark written in front of the "5." As a warm-up for the *E*s, each of them also rated the standardized set of 10 photos. The exact instructions to *E*s were as follows:

Instructions to Experimenters. "You have been asked to participate in a research project developing a test of empathy. You may have seen this project written up in the campus newspaper. There is another reason for your participation in this project—namely, to give you practice in duplicating experimental results. In physics labs, for example, you are asked to repeat experiments to see if your findings agree with those already well established. You will now be asked to run a series of *S*s and obtain from each ratings of photographs. The experimental procedure has been typed out for you and is self-explanatory. DO NOT DISCUSS THIS PROJECT WITH ANYONE until your instructor tells you that you may.

"You will be paid at the rate of $1.00 per hour for your time. If your results come out properly—as expected—you will be paid $2.00 instead of $1.00. The *S*s you are running should average about a —5 rating.

"Just read the instructions to the *S*s. Say *nothing* else to them except hello and goodbye. If for any reason you should say anything to an *S* other than what is written in your instructions, please write down the exact words you used and the situation which forced you to say them.

"GOOD LUCK!"

Results

The results of this experiment are shown in Table II. The difference between the mean photo ratings obtained by *E*s expecting higher ratings and those expecting lower ratings was significant at the .007 level (one-tailed *p*, $t = 3.20$, $df = 8$). Data obtained by the two female *E*s, one in each treatment condition, did not differ from the mean ratings obtained by the male *E*s of their respective experimental conditions. The grades earned by all *E*s in their experimental psychology course were not related to either the mean photo ratings obtained from subjects, or the magnitude of the biasing phenomenon.

Because of the striking nature of the results obtained, we decided to undertake another demonstrational experiment (Rosenthal and Fode, 1963b).

TABLE II
EXPERIMENTAL RESULTS UNDER TWO EXPECTANCIES

Expectancy	
+5	−5
+.66	+.18
+.45	+.17
+.35	+.04
+.31	−.37
+.25	−.42
Means:	
+.40	−.08

Method

Experimenters. Twelve of the 26 male students enrolled in an advanced undergraduate course in industrial psychology were randomly assigned to serve as *E*s. In this sample of *E*s, few were psychology majors; most were majoring in engineering sciences.

Subjects. *S*s were 86 students enrolled in a course in introductory psychology (50 males and 36 females). These *S*s were also given class credit for participating in the experiment. Each *E* ran from 4 to 14 *S*s.

Procedure. The procedure of this experiment was just as before with the exception that *E*s did not handle the photos. Instead, each set of 10 photos was mounted on cardboard and labeled so that *S*s could call out their ratings of each photo to their *E*. It seemed reasonable that less handling of the photos might serve to reduce the effects of *E*s' expectancies on the data obtained from *S*s. As before, half the *E*s were led to expect ratings of +5, and half were led to expect ratings of −5.

Results

The results of this second demonstration are shown in Table III. Once again, all *E*s expecting higher ratings obtained higher ratings than did any *E* expecting lower ratings ($p = .0005$, one-tailed, $t = 4.99$).

In this replication *S*s tended to rate photos as more successful than did the *S*s of the first experiment. This difference was greater for the "+5" groups of *E*s ($p = .004$, two-tailed) than for the "−5" groups ($p = .08$, two-tailed). The second experiment also showed a significantly greater difference than the first between the means of the two treatment conditions ($p = .0005$, two-tailed, $t = 12.25$, $df = 9$). Why the *S*s of the replication rated photos as more successful and why they were signifi-

cantly more biased by their *E*s is difficult to interpret. Any one or more of four fairly obvious reasons may have contributed to these differences:

1. Different samples of *E*s were involved, engineering rather than psychology majors.

2. Different samples of *S*s were involved; but these should not have differed particularly.

3. *E*s did not handle the photos in the second study, as they had in the first.

4. *E*s were instructed by a faculty member in the first experiment and by a graduate student in the replication.

TABLE III

EXPERIMENTAL RESULTS UNDER TWO EXPECTANCIES: REPLICATION

Expectancy	
+5	—5
+3.03	+1.00
+2.76	+0.91
+2.59	+0.75
+2.09	+0.46
+2.06	+0.26
+1.10	—0.49
Means:	
+2.27	+0.48

Conclusions

What can we conclude from the two experiments described? It seems clear from the data that *E*s' expectancies or hypotheses can be partial determinants of the results of their experiments. Since *E*s were not permitted to say anything to their *S*s other than the standard instructions, the communication of *E*'s biases must have been by some subtle paralinguistic (e.g., tone) or kinesic (e.g., facial expressions, gestures) signals.

The *E*s engaged in our studies were not, of course, Ph.D. psychologists. How safe is it therefore to generalize from these student *E*s to "real life" *E*s? Two factors are relevant to the answer. One factor is the increasing amount of data collected in *real life* by other than Ph.D. psychologists. More and more data appear to be collected by research assistants. This trend will likely continue as psychological researchers continue to obtain increasing financial support for their research. Looking now at the pattern for a given psychological researcher, it appears that the better established he has become, the more likely that he has greater financial support for his research. He is therefore able to hire more research assistants. For many of the more productive behavioral researchers, therefore, the actual data collection is often primarily

carried out by research assistants. Most of these assistants are graduate students but there is an increasing number of undergraduate research assistants. This trend is properly endorsed and encouraged by various government programs including the Undergraduate Research Participation Program of the National Science Foundation. It is for these reasons that we feel it to be fairly safe to generalize from our student Es to the real life data collector.

The other factor relevant to the problem of the generalizability of our findings assumes for the moment that *"real"* researchers *are* Ph.D. behavioral scientists. Our position then is that generalizing from advanced students in undergraduate or graduate psychology programs to "real" researchers is somewhat risky business. Risk of this sort, however, is a relative matter and we deem the risk less than the risk of generalizing from the behavior of college freshmen or sophomores to "human behavior" in general, and much less risky than generalizing from infrahuman to human behavior.

IV. Demonstration of Expectancy Effects: Animal Subjects

Pavlov was aware of the fact that experimenters could influence their animal subjects. In speaking of experiments on the inheritance of acquired characteristics, he suggested that the noted increase in learning ability of successive generations of mice was really more an increase in teaching ability on the part of the experimenter (Gruenberg, 1929, p. 327).

Two studies in experimenter expectancy effects have been conducted using animal subjects (Rosenthal and Fode, 1963a; Rosenthal and Lawson, 1963). In the first experiment, a total of 12 experimenters each ran five albino rats on a simple discrimination problem daily for a 5-day period. On the basis of ratings of how well they thought they would like working with rats, six matched pairs were formed. For each pair, one member was assigned to a group of experimenters who were told that the subjects they were running were maze-bright, while the other member was assigned to a group who were told that their subjects were maze-dull. Subjects for the two treatment groups were from a homogeneous colony and matched for age and sex.

Results indicated that on 3 of the 5 days and for the experiment as a whole, experimenters believing their subjects to be bright obtained performance from them significantly superior to that obtained by experimenters believing their subjects to be dull ($p = .01$). The subjects believed to be bright appeared to be learning the problem while those believed to be dull did not. These results occurred in spite of the fact

that on the level of verbal report both groups of experimenters wanted their subjects to perform well. In addition, a research assistant following the identical experimental procedure was able to obtain without "cheating" performance from her animals superior even to that obtained by experimenters believing their subjects were bright. Comparing the degree of correlation between what each experimenter specifically expected to obtain from his animals and what he actually did obtain from them for the "bright" and "dull" groups suggested that these groups were about equally biased, although, of course, in opposite directions.

In the second experiment a total of 38 experimenters were divided into 14 research teams each of which had one rat assigned to it. Eight of the teams were told that the subjects they would be working with had been bred for brightness while the remaining six teams were told that their subjects had been bred for dullness. All subjects were drawn from the same animal colony, all were female, and all were 80 days old. Animals were assigned at random to one of the two treatment conditions, which were experimenters' beliefs or expectations about subjects' ability.

Seven subexperiments including (a) magazine training, (b) operant acquisition, (c) extinction and spontaneous recovery, (d) secondary reinforcement, (e) stimulus discrimination, (f) stimulus generalization, and (g) chaining of responses were performed. Differences in performance favored the groups of experimenters believing their subjects to be bright in seven out of the the eight comparisons and these differences could not often have occurred by chance ($p = .02$). There was no trend over the course of the experiments for the treatment effects to either increase or decrease nor were the performances of subjects in any experiment save one correlated significantly with their performances in the subsequent experiment. It appeared then, that the several experiments were, to a great extent, independent. Comparisons of the treatment effects among each of the five laboratory sections to which experimenters had been assigned showed no real difference, all sections showing the mean differences in the predicted direction and at similar levels of significance.

These differences were obtained in spite of the fact that laboratory instructors gave more help to experimenters whose subjects were performing poorly and that all experimenters were motivated to have their animals perform well in order to complete the sequence of experiments. In addition, a laboratory instructor was present in each laboratory so that gross recording errors and differences in experimenters' treatment of their subjects would have been observed and corrected.

On the basis of questionnaire data obtained in both experiments, it appeared that experimenters believing their subjects to have been bred for brightness were more satisfied with their participation in the experi-

ments, liked their subjects more, watched them more intently, and found them to be more pleasant. They tended also to be more enthusiastic, friendly, encouraging, and pleasant; but were less talkative and less loud when working with their subject. Perhaps the most crucial difference was these experimenters' handling their animals more. The potentially facilitating effect of handling on learning seems well established (Bernstein, 1957).

V. Effects of Excessive Reward on Expectancy Effects

We have seen that experimenters can obtain from their human or animal subjects the data they expect to obtain. The studies described which have employed human subjects offered mild incentives for the operation of the experimenter expectancy effect. Experimenters were offered $2.00 instead of $1.00 "if their data came out as expected." The studies which employed animal subjects offered no incentive for the operation of experimenter biasing of their experimental findings. On the contrary, these experimenters were motivated by the nature of the situation to obtain uniformly good performance from their rat subjects. In these studies then, it appeared that an experimenter's expectancy was a more powerful determinant of the results of his experiment than his desires regarding the results.

In order for any scientist to do any research, he must be motivated to do so, and probably more than casually so. It is, after all, a lot of trouble planning and executing an experiment. The motivation to conduct research is usually related to certain motivations associated with the results of the research. Rarely is the investigator truly disinterested in the results he obtains from his research. William James put it thus: ". . . science would be far less advanced than she is if the passionate desires of individuals to get their own faiths confirmed had been kept out of the game . . . if you want an absolute duffer in an investigation, you must, after all, take the man who has no interest whatever in its results: he is the warranted incapable, the positive fool" (1948, p. 102).

Two experiments were conducted to investigate the effects of "unusual" motivation on the operation of experimenter expectancy effects (Rosenthal, Fode and Vikan-Kline, 1960b; Rosenthal, Friedman, Johnson, Fode, Schill, White, and Vikan-Kline, in press. In the first of these studies, 12 graduate student experimenters ran a total of 58 undergraduate subjects in the photo-rating task described earlier. All experimenters were given an expectancy about the data they would obtain from their subjects. Half the experimenters were moderately motivated to obtain this data by the promise of $2.00 for "obtaining good data." The remaining experimenters were more extremely motivated by the promise

of $5.00 for "obtaining good data." Results indicated that the more moderately motivated experimenters obtained data more in accord with their expectancies than did more extremely motivated experimenters ($p = .01$). There was, in fact, a tendency for the more extremely motivated experimenters to obtain data opposite to that which they had been led to expect.

In the second study 30 advanced undergraduate experimenters ran a total of 150 subjects in the same photo-rating task. Again, all experimenters were given an expectancy about the data they would obtain from their subjects. Half the experimenters were "excessively" motivated by telling them that if the data they obtained was "better" than that of an unknown partner they would be paid not only their own $1.00 but their partner's as well. If, on the other hand, their "partner" did a "better job" he would get their reward as well as his own. Results again suggested that more moderately motivated experimenters obtained data more consistent with their expectancy. Again, more extremely motivated experimenters tended to obtain data opposite to what they had been led to expect. The results of this second study were not so statistically significant, but the two studies taken together yielded a p of .05.

Why might the effect of excessive incentive to bias the subjects' data be to reduce or even to reverse the biasing effect of the experimenters? In a postexperiment group discussion with the experimenters of the second reported study, many of them seemed somewhat upset by the experimental goings-on. Several of them used the term "payola," suggesting that they felt that the investigators were bribing them to get "good" data which was, in a sense, true. Since money had been mentioned and dispensed to only the more motivated experimenters of this study, it seems likely that they were the ones perceiving the situation in this way. Kelman (1953) found that subjects under higher motivation to conform to an experimenter showed less such conformity than did subjects under lower conditions of motivation. One of several of Kelman's interpretations was that the subjects who were rewarded more may have felt more as though they were being bribed to conform for the experimenter's own benefit, thus making subjects suspicious and resentful, and therefore less susceptible to experimenter influence. This interpretation fits the present situation quite well.

VI. Subjects' View of Biased Experimenters

Riecken has discussed the social nature of the experimenter-subject interaction in some detail (1962). An important aspect of this interaction is the subject's perception of the experimenter. What sort of person is the

experimenter seen to be in his relationship to the subject? The answer to this question has general relevance to an increased understanding of the social psychology of the psychological experiment. More particularly, is a more biased experimenter seen as a somewhat different person than a less biased experimenter? The answer to this question might increase our understanding of the mediation of experimenter expectancy effects.

Subjects' Perceptions of Experimenters' Behavior

At the conclusion of one of the experiments involving the photo-rating task each of 56 subjects completed a questionnaire describing the behavior of his experimenter during the conduct of the experiment. Twelve experimenters completed the same forms describing their own behavior during the experiment. Neither experimenters nor subjects knew beforehand that they would be asked to complete these question-naires, and no one save the investigators saw the completed forms (Rosen-thal, Fode, Friedman, and Vikan-Kline, 1960a). These forms consisted of 27 20-point rating scales ranging from —10 (e.g., extremely discourteous) to +10 (e.g., extremely courteous).

Mean ratings of the experimenters by their subjects and by them-selves reflected very favorably on the experimenters. The profile of the experimenters as they were viewed by subjects showed remarkable simi-larity to the profile of the experimenters as viewed by themselves (rho = .89). To summarize and facilitate interpretation of the obtained ratings, all variables were intercorrelated and cluster analyzed (see Table IV). Four clusters emerged for which the associated B coefficients were all considerably larger than generally deemed necessary to establish the sig-nificance of a cluster. For mnemonic purposes we may label Cluster I as "Casual-Pleasant," Cluster II as "Expressive-Friendly," Cluster III as the "Kinesic Cluster," and Cluster IV as "Enthusiastic-Professional."

Experimenters differ in the degree to which their expectancy is predictive of the data they obtain from their subjects. We could therefore ask the question of whether those experimenters whose expectancies de-termined their results more were perceived by subjects as behaving dif-ferently during the course of the experiment. Magnitude of expectancy effect or bias was therefore correlated with all the ratings subjects had made of their experimenters' behavior.

The median correlations with degree of experimenter bias of the variables in Clusters I and IV were .26 and .21, respectively; neither was significantly greater than a correlation having a $p = .50$. The median correlations with degree of experimenter bias of the variables in Clusters II and III were .47 and .43, respectively. Both of these median correla-

tions were significantly greater than a correlation to be often expected by chance; two-tailed p's were .04 and .01, respectively. More biased experimenters, then, were characterized by higher loadings on the "Expressive-Friendly" and the "Kinesic" clusters. These findings suggest that kinesic

TABLE IV

CLUSTER ANALYSIS OF SUBJECTS' PERCEPTION OF EXPERIMENTERS

Cluster I: $B = 6.48$	Cluster II: $B = 3.97$
Honest	Liking
Casual	Friendly
Relaxed	Personal
Pleasant	Interested
Courteous	Encouraging
Business-like	Expressive face
Slow-speaking	Expressive-voiced
Pleasant-voiced	Use of hand gestures
Behaved consistently	Satisfied with experiment
Mean rating = 5.91	*Mean rating = 2.57*
Cluster III: $B = 9.10$	Cluster IV: $B = 3.55$
Use of head gestures	Enthusiastic
Use of arm gestures	Professional
Use of trunk	Quiet (nontalkative)
Use of body	
Use of legs	
Mean rating = —1.96	*Mean rating = 3.01*

and possibly paralinguistic aspects of the experimenter's interaction with his subjects serve to communicate the experimenter's bias to his subjects.

Subjects' Predictions of Experimenters' Computational Errors

All experimenters are rated as very honest by their subjects, but some are rated more so than others. In one of our studies, 11 out of 24 experimenters made computational errors in their data processing in the direction of their expectancy. These experimenters were rated as significantly less honest ($p = .02$) by their subjects during the course of data collection. Just how subjects were able to predict their experimenters' computational errors from their judgments of experimenters' behavior during the experiment is a fascinating question for which we presently have no answer. It seems clear, however, that subjects learn a good deal about their experimenter in the brief interaction of the person-perception experiment conducted. Furthermore, when a subject rates the behavior of an experimenter, we may do well to take his rating seriously. It may prove to be a valuable predictive, postdictive, or paridictive variable (Rosenthal *et al.*, in press).

VII. Interaction of Subjects' and Experimenters' Expectancies

In the experiments described so far, and those to be reported here, experimenters' expectancies have been experimentally manipulated. In this section we shall describe the additional effects of the simultaneous experimental manipulation of subjects' expectancies.

Riecken (1962) has discussed the subjects' "deutero-problem" in the psychological experiment, which exists beyond the problems of simply performing the formal task required of him by the experimenter. The most favorable solution to the deutero-problem would involve the maximal satisfaction of the subject's three aims in the experiment: (1) the attainment of some sort of reward, (2) the divination of the experimenter's true purposes, and (3) the favorable presentation of self for the experimenter's scrutiny. Those stimuli in the experimental situation which provide cues to the subject as to how he may best solve his deutero-problem have been called "demand characteristics" (Orne, 1962). One of the major purposes of the two studies to be reported here was to investigate the role of subtle and of blatant demand characteristics in determining the results of psychological research (Rosenthal, Persinger, Vikan-Kline, and Fode, 1963d; White, 1962).

In the first of these studies, 18 experimenters positively reinforced a total of 65 subjects for high positive ratings of the success of persons pictured on photographs. Half the experimenters were led to expect high rates of awareness from their subjects, while the remaining experimenters were led to expect low rates of awareness. Each of these two groups was further divided into a group of experimenters whose subjects' instructions favored their "seeing through" the experimental situation and a group whose subjects received standard instructions. An additional group of four experimenters running a total of 26 standard-instruction subjects was used as a control group. These experimenters were biased to expect high photo ratings but did not reinforce any subject's response.

Experimenters biased to expect higher rates of awareness obtained higher rates of awareness ($p = .07$) and subjects who had been given a set to "see through" the experiment tended to be more often aware, though this effect was not as reliable as the effect of experimenter's bias. The obtained effects were most powerful and most significant statistically ($p = .05$) when they were operating conjointly. Subject set was, however, significantly related to magnitude of conditioning with subjects set to "see through" conditioning less. Sets favoring "seeing through" an experiment may carry an implied set to not conform to more superficial demand characteristics of the experimental situation.

In the second experiment, 18 experimenters ran a total of 108 sub-

jects in the photo-rating task described earlier. Six different expectancies (ranging from −6 to +6) were induced among both experimenters and subjects. The expectancies created in subjects were far less subtle than was the case in the preceding study. Results suggested the operation of contrast effects wherein subjects confronted with experimenters of radically different expectancies seemed to make more entrenched responses. In addition to this contrast effect and regardless of the direction of the expectancies involved, subjects tended to rate the photos as more successful when their expectancy and that of their experimenter were in greater accord.

These studies served to increase the generality of the findings bearing on the effects of experimenters' and subjects' expectancies. In addition, however, they indicated that more subtle demand characteristics had more predictable effects on the responses made by subjects. As with unusually motivated experimenters, unusually "expectant" or cued subjects provided data less directly predictable from a knowledge of experimenters' hypotheses.

VIII. Effects of Experimenter Expectancy: Negative Instances

The title of this section may prove misleading. It may imply that until the work reported here was accomplished, there was only unequivocal and highly predictable and predicted evidence bearing on the question of the effects of experimenters' expectancies on the results of their research. Nothing could be further from the truth. The results of the first experiment described in the section on the effects of excessive reward were at least in part quite unpredicted. *Post hoc*, these results made sense but it was, after all, *post hoc*. With the results of the second experiment reported in that section we gained confidence that our initially "negative" instances were negative only with respect to our initial predictions. They made sense after further data collection and seemed consistent with each other.

White's (1962) experimental data, summarized in the last section, similarly did not conform to our expectation. There was no neat regression line ascending from congruent experimenter-subject expectations for negative data to those expectations congruent for positive data. In one sense, that study also provided a negative instance. More data are needed to make conceptual good sense of a hopefully orderly phenomenon. As we turn then to a consideration of "negative" instances, we must bear in mind that they are neither the first nor the last such instances. Nor does their "negative" nature imply that we cannot learn from them something useful about the phenomenon under discussion.

In the first of the two studies to be summarized here, eight experimenters administered an intelligence test (Block design of the WAIS) to to a total of 32 subjects (Wartenberg-Ekren, 1962). Experimenters were to be led to expect that half their subjects would perform well and half would perform poorly.

Results of this study showed no effect of experimenters' expectancies on the data obtained from subjects. Wartenberg-Ekren's analysis of experimenters' questionnaires revealed that all experimenters expected the "brighter" subjects to perform better. In addition, none of the experimenters felt that they had treated their subjects differentially as a function of their perceived group membership. The over-all negative findings may have been due in part to the fact that one of the eight experimenters intentionally instituted a double blind procedure to avoid bias! He simply did not look at the code identifying the treatment condition of his next subject. Another experimenter was "blind" for half his subjects, although not intentionally so.

Subjects' ratings of their experimenters in the Wartenberg-Ekren study were not directly comparable to those of our earlier studies because she used a three-point scale rather than our 20-point scale. We were able, however, to group many of her scales into two of our four earlier derived clusters: Cluster I (Casual-Pleasant) and Cluster II (Expressive-Friendly). Subjects labeled as brighter did not rate their experimenters any differently on the scales included in Cluster I. These subjects, however, rated their experimenters as significantly higher on Cluster II variables than did subjects labeled as less bright ($p = .03$). This suggests that although all experimenters believed they treated all subjects alike, they made some distinction in their treatment of subjects as a function of their perceived group membership. We should emphasize that not only were subjects unaware of which group they belonged to, indeed, they were unaware that they belonged to *any* group. It seems safe to accept their ratings of their experimenter as reflecting real differences in experimenters' treatment of subjects. The Cluster II variables included in the Wartenberg-Ekren study, and on which "brighter" subjects rated their experiments as higher, were friendly, likeable, interested, encouraging, expressive-faced, and use of hand gestures. Apparently even the most careful conscious safeguards against differential treatment of subjects in different treatment conditions may prove ineffective.

The second study to be summarized in this section is one conducted by Pflugrath (1962). He had three sets of three graduate student counselors administer the Taylor Manifest Anxiety Scale to a total of 142 subjects. One set of experimenters was led to believe their groups of subjects were very anxious, another set was led to believe their groups of

subjects were nonanxious, and the third set was told nothing about their groups of subjects.

The over-all analysis of the results showed no significant differences in anxiety scores earned by the subjects of the three groups of experimenters. Among the subjects run under the condition of their experimenter believing them to be highly anxious, a significantly greater proportion scored as relatively less anxious than the subjects of the control group ($p = .02$). This finding was unpredicted. Nevertheless, the phenomenology of the situation must be considered. Here we have a group of counselors-in-training told nothing about their subjects—the control group. Another group of counselors is told their subjects are not anxious—so much the better for the subjects! But when counselors are told they will be testing very anxious subjects who even have had contact with the counseling center, might not all their counseling skills be subtly brought to bear upon the task of reducing their subjects' anxiety? This interpretation seems so plausible, at least on a *post hoc* basis, that we wonder why we did not predict it in the first place.

The simplest conclusions to draw from the results of the experiments discussed in this section might be that experimenters' expectancies do not affect subjects' intelligence test or anxiety scale performance. Indeed, there may be a large array of tasks which will prove relatively resistant to the effects of experimenter expectancies. What sorts of tasks are more or less resistant is a question deserving further research. Some questions remain, however, about the particular two experiments summarized. To what extent was the absence of an expectancy effect in the intelligence test study due to the intentional and unintentional instituting of a double-blind procedure, which if it remains truly blind, must eliminate expectancy effects?

Less in the nature of a question and more as an illustration of the potential power of expectancy effects, we have the subjects' word for it that experimenters treated them differentially as a function of the experimenters' expectancies, the existence of which were unknown to the subjects, and this in a context of tremendous emphasis on experimenters' need for objectivity, standardization, and nondifferential treatment of subjects.

Whether therapeutically oriented experimenters in the social sciences generally try to reduce the anxiety of those subjects perceived to be anxious is a question deserving of further research. Not only in administering anxiety scales, but in a great deal of contemporary social research, subjects are exposed to conditions believed to make them anxious. What might be the effect on the outcome of experiments of this sort of the covert therapeutic zeal and/or skill of various investigators

carrying out this research? Might certain investigators typically conclude "no difference" because they tend to dilute the effects of treatment conditions? And conversely, might others be led to conclude "significant difference" by their increasing the anxiety of subjects known to belong to the "more anxious condition" of an experiment? Clearly, these are not necessarily effects of experimenters' expectancies, but they are effects of experimenter attributes which may have equally serious implications for how we do research in the social sciences.

IX. The Effect of Early Data Returns

We have been discussing the effects of experimenters' expectancies on the data they obtain from their subjects. We shall discuss here the effect of the data experimenters obtain upon their expectancies. Except in the most exploratory or heuristic experimental enterprise, the experimenter's expectancies are likely to be based upon some sort of data. It need not be formal, experimental data. It may be based upon a nearly casual observation of behavior. The observation need not even be the experimenter's own. If data are in any case the most likely determinant of experimenter expectancies, then we may fairly ask: What about the data obtained early in an experiment? What are their effects upon data subsequently obtained within the same experiment?

That the "early returns" of psychological research studies can have an effect on experimenters' expectancies was noted and well discussed by Ebbinghaus (1885). After saying that investigators notice the results of their studies as they progress he stated, "Consequently it it unavoidable that, after the observation of the numerical results, suppositions should arise as to general principles which are concealed in them and which occasionally give hints as to their presence. As the investigations are carried further, these suppositions, as well as those present at the beginning, constitute a complicating factor which probably has a definite influence upon the subsequent results" (p. 28). He went on to speak of the pleasure of finding expected data and surprise at obtaining unexpected data and continued by stating the hypothesis of the present studies: Where "average values" were obtained initially, subsequent data would tend to be less extreme and where "especially large or small numbers are expected it would tend to further increase or decrease the values" (p. 29).

Ebbinghaus was, of course, speaking of himself as both experimenter and subject. Nevertheless, on the basis of his thinking and of the reasoning described earlier, it was decided to test Ebbinghaus's hypothesis of the effect of early data returns on data subsequently obtained by experimenters.

Two experiments were conducted. In the first of these studies (Rosenthal *et al.*, 1963c) 12 biased experimenters, each running six subjects on the photo-rating task, were equally and randomly divided into three treatment conditions. One group of experimenters obtained "good" or expected data from their first two subjects (who were actually accomplices), another group of experimenters obtained "bad" or unexpected data from their first two subjects (who were also accomplices), while the third group, utilizing only naïve subjects, served as a control. Comparisons were made of the mean data obtained by experimenters from the last four naïve subjects run.

Results indicated that experimenters obtaining "good" initial data also obtained good subsequent data. Experimenters obtaining "bad" initial data obtained bad subsequent data.

In the second study (Rosenthal, Kohn, Marks, and Carota, 1963a), 26 experimenters each running about six subjects were given one of two opposite experimental expectancies. By the same method employed in the earlier study, half the experimenters had their expectancies confirmed by their first two subjects and half had their expectancies disconfirmed. Within each of the conditions described, half the experimenters were praised for their experimental technique and half were reproved. Results of this study were more complex than those of the first experiment, but showed that experimenter expectancies were most clearly determinants of experimental results when they were confirmd by the early data returns. This effect depended, however, on the personality of the investigator who administered the praise or reproof to the experimenter.

X. The Mediation of Experimenter Expectancy Effects

Granting the occurrence and some generality of experimenter expectancy effects, how do they operate? Cheating cannot reasonably account for the observed effects since at least those instances of cheating of which we have become aware tended on the whole to diminish the biasing effect, as when experimenters who believed their rat subjects to be dull prodded them, and, in a very few cases, presented fraudulent data. Experimenters' data recording and computations were checked and were found to be so accurate that, in general, written recording and computational errors could not be reasonably implicated as agents of expectancy effects. If neither cheating nor honest errors could account for our findings, what might?

A. Verbal Conditioning

The most obvious hypothesis seemed to be some form of verbal conditioning. If an experimenter expects to obtain high ratings of photographs, might he not subtly reinforce this type of response? Conversely,

if he expects low ratings of photos, might he not reinforce subtly those responses which are low? He might be capable of this system of subtle reinforcement even without any implication of dishonesty, for it might be an unintended response on his part. Fortunately, we were able to test this hypothesis. If indeed verbal conditioning were mediating the phenomenon, we might expect to find that expectancy effects increase as a function of the number of photos rated. Certainly, we would not expect to find any biasing on the very first photos rated by subjects run by different groups of experimenters. There had, after all, been no reinforcement possible prior to the very first response. In a test of this hypothesis (Rosenthal et al., 1964) we found, if anything, that biasing decreased over the course of the photo ratings. Furthermore, there was a significant biasing effect in evidence on the first photo alone, thus ruling out verbal conditioning as a necessary mediator or even as an augmentor of the phenomenon. An important implication of this finding is our need to pay special attention in our search for the mediators of the expectancy bias phenomenon to the brief pre-data-gathering interaction during which the experimenter greets, seats, "sets," and instructs the subject.

A subsequent study (Fode, Rosenthal, Vikan-Kline, and Persinger, 1961) was conducted to learn whether operant conditioning *could* drive the ratings of photos up or down according to the will of the experimenter. Results showed clearly that this was possible, and that it worked best with certain types of subjects. We may therefore conclude that, while verbal conditioning is neither a necessary nor necessarily frequent antecedent of biasing, it nevertheless could be.

B. MODALITY OF CUE COMMUNICATION

Are visual or verbal cues, such as tone, more important for the mediation of bias? Fode (1960) studied this question by using a group of experimenters behind screens to eliminate visual cues, and a group of experimenters who remained silent throughout the experiment to eliminate verbal cues. He found that verbal cues of tone are probably sufficient to mediate expectancy effects but that they can be greatly augmented by visual cues. Restriction of visual cues accounted for about 80% of the variance of bias magnitude.

C. PERSONAL CHARACTERISTICS OF EXPERIMENTERS AND SUBJECTS

It seems reasonable to suppose that it is the behavior of the experimenter in the experimental situation that determines the occurrence and magnitude of expectancy effects. Without knowing exactly what occurs in the experimenter-subject interaction we can only posit that certain personal characteristics of experimenters are likely to be associated with different behavior vis-à-vis the subject in the experimental situation. As

a first step, therefore, we have tried to find some more enduring personal characteristics associated with greater or lesser experimenter expectancy effects.

Experimenter's need for approval as measured primarily by the Marlow-Crowne Social Desirability Scale (M-C SD) appears to predict expectancy effects in a fairly reliable way. In seven samples (Persinger, 1962; Rosenthal, 1963d; Rosenthal et al., 1963b), experimenters' need for approval was correlated with the magnitude of their expectancy effects. For a total of 57 experimenters of medium anxiety level (or unselected for anxiety) and previously not acquainted with their subjects, the obtained correlations averaged .64 ($p < .001$). A recently completed study, however, yielded contradictory data (Rosenthal et al., 1963a). In that study experimenters with higher need for approval scores showed less expectancy effects ($p = .04$). Disregarding that finding for the moment, it appears that experimenters who may have a greater need to please the principal investigators obtain more of the sort of data they are led by him to expect; at least when their anxiety level is average. For a total of 50 experimenters scoring either very high or very low on the Taylor Manifest Anxiety Scale (MAS) the correlations between need for approval and expectancy effects tended to be negative though not significantly so. In none of our studies have we obtained a significant relationship between subjects' need for approval and their susceptibility to experimenter expectancy effects.

Experimenter's anxiety level (MAS) has been found related to expectancy effect in a remarkably inconsistent manner. In three experiments, experimenters with medium levels of anxiety showed the greatest expectancy bias (Fode, 1963; Rosenthal et al., 1963a; Rosenthal et al., 1963b). In two experiments, experimenters with high levels of anxiety showed the greatest expectancy bias (Rosenthal et al., 1962; Rosenthal et al., 1963e). Finally one experiment showed experimenters scoring low in anxiety to have the greatest expectancy bias (Persinger, 1962). This confusion of findings is matched by the irregularity of the correlation between subjects' level of anxiety and their susceptibility to experimenter expectancy effects. In two experiments medium anxious subjects were most susceptible (Fode, 1963; Vikan-Kline, 1962). In two other studies highly anxious subjects were most susceptible (Rosenthal et al., 1963e; Rosenthal et al., 1963b). In a single study, least anxious subjects were found most susceptible (Persinger, 1962) and in another, medium anxious subjects were found least susceptible (Rosenthal et al., 1963a). We can only conclude that experimenters' and subjects' anxiety level are related significantly to expectancy effects in an as yet indeterminate manner.

In most of our studies we have employed primarily male experi-

menters. In three experiments enough female experimenters were included to allow us to compare the magnitude of expectancy effects for male and female data collectors. In all three cases male experimenters showed the greater expectancy effects (Marcia, 1961; Persinger, 1962; Rosenthal et al., 1963b). It may be that the culturally prescribed relative assertiveness of males may make them more successful sources of unintended social influence. Again on the basis of role prescription we might expect female subjects to be more susceptible to the unintended social influence process implied by expectancy effects in the experimental situation. In several studies we found no difference between male and female subjects in this regard. In three experiments where significant differences were obtained, however, we did find female subjects more susceptible to expectancy effects (Persinger, 1962; Rosenthal et al., 1963b; Rosenthal et al., 1963c). These findings were heartening in their consistency within themselves and with the general, analogous findings in the more usual literature on susceptibility to interpersonal influence processes.

Prior acquaintanceship between male experimenters and their subjects seemed to facilitate the expectancy biasing phenomenon. In one experiment this appeared to be true for both male and female subjects (Rosenthal et al., 1963e), but in a second study (Persinger, 1962) it appeared true only for female subjects. The perceived status of the experimenter was also found to be related to his degree of biasing. Vikan-Kline (1962) found higher-status experimenters better able to bias their subjects' responses.

If, on the basis of the data available to date, we were forced to describe the paradigm fostering maximal experimenter expectancy effects, we would postulate an experimenter with a high need for social approval and with an anxiety level neither very high nor very low. The experimenter would have high status, be gesturally inclined, and behave in a friendly, interested manner vis-à-vis his subjects. Subjects might best be acquainted with their experimenter and be female rather than male.

The pattern described might be understood best by considering the experimental dyad as a signal exchange system. The signals under discussion are, of course, unintentional. Experimenters high in need for social approval may typically be more precise in their signaling behavior. The business of impression management (Goffman, 1956), or signal editing, is more important to them in their everyday life. Their motivation for biasing may also be greater because of their need to please the source of the hypothesis. The high status and friendly manner may serve to focus subjects' attention onto the signal source and increase the likelihood of the experimenter's unintentional message being understood.

XI. Methodological Implications of Studies of Expectancy Effects

It seemed reasonable to conclude from our findings that systematic expectancy effects might be eliminated by employing as data collectors, research assistants who did not know the experimenters' hypothesis or expectancy. Not only did this technique seem logically implied by our data, but it would be practical as well. More and more data collection is actually carried out by research assistants. We decided, however, to test the soundness of this methodological suggestion (Rosenthal *et al.*, 1963e).

We began by conducting a by now fairly standard experiment in experimenter expectancy effects. Fourteen experimenters ran a total of 76 subjects in the photo-rating task with half the experimenters led to expect +5, and half led to expect −5 mean ratings of the success of persons pictured in photos. At the conclusion of this experiment, each experimenter was awarded a "research grant" from which he could draw a small salary and also hire two research assistants. Assistants were randomly assigned to experimenters who then trained and supervised their two assistants. Each of these then ran 5 or 6 randomly assigned subjects of his own. Unlike the original instructions to experimenters, instructions to their assistants did not inform them as to what perceptions to expect from their subjects. Experimenters, however, were subtly led, by their printed instructions, to expect their assistants to obtain data of the same sort they had themselves obtained from their own subjects.

Experimenters biased their subjects, and assistants in turn biased *their* subjects. The correlation between magnitude of experimenters' bias and their respective pair of assistants' bias was .67 ($p = .01$). Apparently, the experimenters' hypothesis or expectancy may be communicated to his research assistants without the assistants being told the nature of the hypothesis or expectancy.

What methodological suggestions remain then which might serve to reduce or eliminate experimenter expectancy effects? For those studies in which it is possible to do so, the experimenter might eliminate himself and his surrogates from the interaction with subjects. Automated setups make this feasible for some kinds of behavioral research, but not for others. Any technique of instruction of data collectors by the principal investigator which would eliminate the possibility of the subtle communication of his expectancies would be a methodological improvement. This would be no easy matter and no perfect solution. The too frequent failure of the double-blind method in medical research attests to this. It is a failure not of "double-blindness" but of maintenance of "blindness." During the experimenter-subject interaction each may learn too much about the other to insure "blindness-maintenance."

Not only because of the danger of bias, but also because of the general nature of experimenter differences, it would be desirable to employ samples of experimenters drawn as randomly as possible from a relevant population of relevantly uninformed experimenters. Following Brunswik (1956), this would greatly increase the generality of all our research findings and thus be of benefit even if *no* bias were ever operating. Alternatively, there may be value in employing samples of experimenters with known distributions of bias as Mosteller (1944) has suggested in the case of interviewers. The particular biases, however, need not be pre-existing ones and it may be useful to *purposefully* induce different biases in our sample of experimenters, giving us better control over the nature and degree of experimenters' biases. That this is a practical technique has been demonstrated recently by Rosenhan (personal communication), who had his own research repeated by an experimenter in whom the opposite hypothesis had been ingeniously induced.

One fairly concrete implication for the design of experiments may emerge from our research, the addition of "expectancy control groups." In any study employing an experimental (treatment) and a control (no treatment) condition, a group would be added for whom the experimenter(s) is reasonably led to expect the same sort of data as is expected from the treatment group but in which the treatment is not administered. Differences between the treatment group and the expectancy control group might then be attributable to the treatment truly or to a treatment and expectancy interaction rather than to expectancy alone.

XII. Some Substantive Implications of Studies of Expectancy Effects

Perhaps the most compelling and the most general conclusion to be drawn is that human beings can engage in highly effective and influential unprogrammed and unintended communication with one another. The subtlety of this communication is such that casual observation of human dyads is unlikely to reveal the nature of this communication process. Sound motion pictures may provide the necessary opportunity for more leisurely, intensive, and repeated study of subtle influential communication processes. We have obtained sound motion picture records of 29 experimenters each interacting with several subjects. Preliminary analyses have given us cause to hope that we may be able to learn something of consequence about the mediation of expectancy effects in particular, and about subtle communication processes in general. In these films, all experimenters read identical words to their subjects so that the burden of communication falls on the gestures, expressions, and intonations which

accompany the highly programmed aspects of experimenters' inputs into the experimenter-subject interaction.

Some interesting practical questions arise from these considerations. When an experienced physician or psychotherapist tells the neophyte therapist that the neophyte's patient has a poor or good prognosis, is the experienced clinician only assessing or is he actually "causing" the poor or good prognosis? When the master teacher tells his apprentice that a pupil appears to be a slow learner, is this prophecy then self-fulfilled? When the employer tells the employee that a task cannot be accomplished, has the likelihood of its accomplishment thereby been reduced? More subtly, might these phenomena occur even if the supervisors never verbalized their beliefs? The data cited suggest that they may.

References

Azrin, N. H., Holz, W., Ulrich, R., and Goldiamond, I. The control of the content of conversation through reinforcement. *J. exp. anal. Behav.*, 1961, 4, 25-30.

Bernstein, L. The effects of variations in handling upon learning and retention. *J. comp. physiol. Psychol.*, 1957, 50, 162-167.

Boring, E. G. *A history of experimental psychology.* (2nd ed.) New York: Appleton-Century-Crofts, 1950.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. of California Press, 1956.

Cahalan, D., Tamulonis, Valerie, and Verner, Helen W. Interviewer bias involved in certain types of opinion survey questions. *Int. J. opin. attit. Res.*, 1947, 1, 63-67.

Crespi, L. P. The cheater problem in polling. *Publ. Opin. Quart.*, 1945-1946, 9, 431-435.

Ebbinghaus, H. *Memory: a contribution to experimental psychology.* (1885) Translated by Ruger, H. A., and Bussenius, Clara E. New York: Teachers College, Columbia Univer., 1913.

Fisher, R. A. Has Mendel's work been rediscovered? *Ann. Sci.*, 1936, 1, 115-137.

Fode, K. L. The effect of non-visual and non-verbal interaction on experimenter bias. Unpublished master's thesis, Univer. of North Dakota, Grand Forks, 1960.

Fode, K. L. The effect of experimenters' and subjects' level of anxiety and need for approval on experimenter outcome-bias. Unpublished doctoral dissertation, Univer. of North Dakota, Grand Forks, 1963.

Fode, K. L., Rosenthal, R., Vikan-Kline, Linda L., and Persinger, G. W. Susceptibility to influence in a verbal conditioning situation. Unpublished data, Univer. of North Dakota, Grand Forks, 1961.

Friedman, Pearl. A second experiment on interviewer bias. *Sociometry*, 1942, 5, 378-379.

George, W. H. *The scientist in action: a scientific study of his methods.* New York: Emerson Books, 1938.

Goffman, E. *The presentation of self in everyday life.* Univer. Edinburgh Soc. Sci. Res. Centre, Monogr. 2, 1956.

Gruenberg, B. C. *The story of evolution.* Princeton, New Jersey: Van Nostrand, 1929.

Hanley, C., and Rokeach, M. Care and carelessness in psychology. *Psychol. Bull.*, 1956, 53, 183-186.

James, W. *Essays in pragmatism.* New York: Hafner, 1948.

Kelman, H. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, 6, 185-214.

Lindzey, G. A note on interviewer bias. *J. appl. Psychol.*, 1951, 35, 182-184.

Mahalanobis, P. C. Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. statist. Soc.*, 1946, 109, 325-370.

Marcia, J. Hypothesis-making, need for social approval, and their effects on unconscious experimenter bias. Unpublished master's thesis, Ohio State Univer., Columbus, Ohio, 1961.

Merton, R. K. The self-fulfilling prophecy. *Antioch Rev.*, 1948, 8, 193-210.

Mosteller, F. Correcting for interviewer bias. In H. Cantril, (Ed.), *Gauging public opinion*. Princeton, New Jersey: Princeton Univ. Press, 1944, pp. 286-288.

Orne, M. T. On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *Amer. Psychologist*, 1962, 17, 776-783.

Pearson, K. On the mathematical theory of errors of judgment wtih special reference to the personal equation. *Phil. Trans. Roy. Soc. Lond.*, 1902, 198, 235-299.

Persinger, G. W. The effect of acquaintanceship on the mediation of experimenter bias. Unpublished master's thesis, Univer. of North Dakota, Grand Forks, 1962.

Pflugrath, J. Examiner influence in a group testing situation with particular reference to examiner bias. Unpublished master's thesis, Univer. of North Dakota, Grand Forks, 1962.

Pfungst, O. *Clever Hans (the horse of Mr. von Osten); a contribution to experimental, animal, and human psychology*. Translated by Rahn, C. L. New York: Holt, 1911.

Riecken, H. A program for research on experiments in social psychology. In N. F. Washburne, (Ed.), *Decisions, values and groups*. New York: Pergamon Press, 1962, Vol. II, pp. 25-41.

Roe, Anne. The psychology of the scientist. *Science*, 1961, 134, 456-459.

Rosenthal, R. The effects of the experimenter as observer on the results of the psychological experiment. Unpublished manuscript, Harvard Univer., Cambridge, Massachusetts, 1963. (a)

Rosenthal, R. Experimenter attributes as determinants of subjects' responses. *J. proj. Tech. Pers. Assess.*, 1963, 27, 324-331. (b)

Rosenthal, R. Experimenter modeling effects as determinants of subjects' responses. *J. proj. Tech. Pers. Assess.*, 1963, 27, 467-471. (c)

Rosenthal, R. On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. *Amer. Scient.*, 1963, 51, 268-283. (d)

Rosenthal, R. Experimenter outcome-orientation and the results of the psychological experiment. *Psychol. Bull.*, in press. (a)

Rosenthal, R. The volunteer subject. *Hum. Relat.*, in press. (b)

Rosenthal, R., and Fode, K. L. The effect of experimenter bias on the performance of the albino rat. *Behav. Sci.*, 1963, 8, 183-189. (a)

Rosenthal, R., and Fode, K. L. (Psychology of the scientist: V.) Three experiments in experimenter bias. *Psychol. Rep. Monogr.*, 3, V12, 1963. (b)

Rosenthal, R., and Lawson, R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. *Amer. Psychologist*, 1963, 18, 345. (Abstr.)

Rosenthal, R., Fode, K. L., Friedman, C. J., and Vikan-Kline, Linda L. Subjects' perception of their experimenter under conditions of experimenter bias. *Percept. mot. Skills*, 1960, 11, 325-331. (a)

Rosenthal, R., Fode, K. L.. and Vikan-Kline, Linda. The effect on experimenter bias of varying levels of motivation of Es and Ss. Unpublished manuscript, Harvard Univer., Cambridge, Massachusetts, 1960. (b)

Rosenthal, R., Persinger, G. W., and Fode, K. L. Experimenter bias, anxiety, and social desirability. *Percept. mot. Skills*, 1962, 15, 73-74.

Rosenthal, R., Kohn, P., Marks, Patricia, and Carota, N. Experimenters' hypothesis-confirmation and mood as determinants of experimental results. Unpublished manuscript, Harvard Univer., Cambridge, Massachusetts, 1963. (a)

Rosenthal, R., Persinger, G. W., Mulry, R. C., Vikan-Kline, Linda, and Grothe, M. A motion picture study of 25 biased experimenters. Unpublished manuscript, Harvard Univer., Cambridge, Massachusetts, 1963. (b)

Rosenthal, R., Persinger, G. W., Vikan-Kline, Linda, and Fode, K. L. The effect of early data returns on data subsequently obtained by outcome-biased experimenters. *Sociometry*, 1963, 26, 487-498. (c)

Rosenthal, R., Persinger, G. W., Vikan-Kline, Linda, and Fode, K. L. The effect of experimenter outcome-bias and subject set on awareness in verbal conditioning experiments. *J. verb. Learn. verb. Behav.*, 1963, 2, 275-283. (d)

Rosenthal, R., Persinger, G. W., Vikan-Kline, Linda, and Mulry, R. C. The role of the research assistant in the mediation of experimenter bias. *J. Pers.*, 1963, 31, 313-335. (e)

Rosenthal, R., Fode, K. L., Vikan-Kline, Linda, and Persinger, G. W. Verbal conditioning: mediator of experimenter expectancy effects? *Psychol. Rep.*, 1964, 14, 71-74.

Rosenthal, R., Friedman, C. J., Johnson, C. A., Fode, K. L., Schill, T. R., White, C. R., and Vikan-Kline, Linda. Variables affecting experimenter bias in a group situation. *Genet. Psychol. Monogr.*, in press.

Shapiro, A. P. The investigator himself. In S. O. Waife and A. P. Shapiro (Eds.), *The clinical evaluation of new drugs.* New York: Hoeber-Harper, 1959.

Sheffield, F. D., Kaufman, R. S., and Rhine, J. B. A PK experiment at Yale starts a controversy. *J. Amer. Soc. Psychical Res.*, 1952, 46, 111-117.

Stanton, F. Further contributions at the twentieth anniversary of the psychological corporation and to honor its founder, James McKeen Cattel. *J. appl. Psychol.*, 1942, 26, 16-17.

Stanton, F., and Baker, K. H. Interviewer bias and the recall of incompletely learned materials. *Sociometry*, 1942, 5, 123-134.

Symposium: Survey on problems of interviewer cheating. *Int. J. opin. attit. Res.*, 1947, 1, 93-106.

Turner, J. What laymen can ask of scientists. *Science*, 1961, 133, 1195.

Verplanck, W. S. The control of the content of conversation: reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, 51, 668-676.

Vikan-Kline, Linda. The effect of an experimenter's perceived status on the mediation of experimenter bias. Unpublished master's thesis, Univer. of North Dakota, Grand Forks, 1962.

Wartenberg-Ekren, Ursula. The effect of experimenter knowledge of a subject's scholastic standing on the performance of a reasoning task. Unpublished master's thesis, Marquette Univer., 1962.

White, C. R. The effect of induced subject expectations on the experimenter bias situation. Unpublished doctoral dissertation, Univer. of North Dakota, Grand Forks, 1962.

Wilson, E. B. *An introduction to scientific research.* New York: McGraw-Hill, 1952.

Winkel, G. H., and Sarason, I. G. Subject, experimenter, and situational variables in research on anxiety. *J. abnorm. soc. Psychol.*, in press.

Wolins, L. Responsibility for raw data. *Amer. Psychologist*, 1962, 17, 657-658.

Yule, G. U. On reading a scale. *J. Roy. statist. Soc.*, 1927, 90, 570-587.

PERSONALITY AND CONDITIONING

S. H. Lovibond

UNIVERSITY OF ADELAIDE, ADELAIDE, SOUTH AUSTRALIA[1]

I. Introduction

Since the publication of the early work of Pavlov (1927), which described marked individual differences in the conditioning performance of animals, the method of conditioning has held promise of elucidating basic parameters of human personality functioning. Early attempts to apply the method to human subjects encountered serious methodological difficulties. Apart from such technical problems as the measurement of the output of saliva in salivary conditioning of human subjects, it became apparent that the central factors of "attitude," "set," or "expectancy" are important sources of variability in human conditioning. The conditioning of other responses, such as the GSR, brought new problems of measurement. It is probably true, however, that the most significant barriers to progress were conceptual, rather than methodological. In the Soviet Union there was apparently a long period in which little advance

was made on Pavlov's thinking in relation to the problems of individual differences in conditioning. In the Western world this aspect of Pavlov's theorizing had little impact, and there were no alternative theories which might have systematized the findings of existing research, and pointed the way to significant areas of future study.

The resurgence of Western interest in problems of conditioning and personality over the past decade or so seems to have resulted chiefly from the development of two psychological theories of relevance to problems of personality. These are the theories of Hull (1943, 1952) and Eysenck (1957). Between them these two theories have inspired the majority of Western studies of conditioning in relation to personality functioning over the past decade. Most of the remaining Western investigations have been largely empirical in character. In the USSR, interest continues in the problem of developing Pavlov's typological conceptions.

A review of recent research suggests that studies of personality dynamics through conditioning may most conveniently be grouped under two main headings. First, there are the studies which begin with groups differentiated on the basis of personality characteristics, and proceed to examine the relationship between these personality characteristics and parameters of conditioning performance. Possibly the most important subgroupings within this class of investigation may be defined in terms of the theoretical bases from which the studies derive. Thus the study may be designed to test predictions from a theory of personality of wide generality, it may be concerned with predictions derived from a theory of limited aspects of personality functioning, or it may be unrelated to systematic theory of any sort.

In the second group of studies, techniques of conditioning are used in the process of defining major personality variables. Virtually all investigations of personality and conditioning in Western countries fall into the first major class, whereas the most significant studies in the USSR belong to the second major grouping.

A further development leading to renewed interest in personality–conditioning relations has been the application of direct conditioning methods in the therapy of behavior disturbances. This work received a good deal of its impetus from the application of conditioning methods in the study of experimental neuroses. Whereas the studies previously considered are concerned with understanding personality dynamics, investigations of experimental neuroses and behavior therapy are concerned with changing certain aspects of personality functioning.

An examination of the literature on personality and conditioning quickly reveals the extent to which further advance is dependent as much

on the solution of basic issues in conditioning as on conceptual development in relation to personality functioning. For this reason it has been found necessary to devote a good deal of attention in the following pages to problems of methodology, and to such fundamental issues as the criteria of conditioning.

Although the procedures of classical conditioning of involuntary responses are often referred to as "simple conditioning," it is becoming increasingly clear that, at this level of behavioral modification, we are confronted with sufficient complexity to make interpretation far from straightforward. Since the study of the instrumental conditioning of voluntary responses in human beings (e.g., verbal conditioning) introduces further complexity, it has been considered wisest to place exclusive emphasis on classical procedures at this stage. In any case, classical conditioning studies account for the great majority of the investigations of personality and conditioning which have been carried out to date.

Recently Diamond, Balvin, and Diamond (1963) have declared in favor of a return to the use of the terms "conditional" and "unconditional" in place of "conditioned" and "unconditioned" reflexes. These authors point out that "the context of Pavlov's discussion never leaves any doubt that the meaning he wished to convey is that of the words conditional and unconditional. He spoke of the conditional reflexes as 'temporary connections,' and he emphasized that they lack the certainty and regularity of occurrence of innate reflexes which can, by contrast, be called unconditional; that is, the reaction is sure to occur in response to the adequate stimulus despite other circumstances and conditions" (Diamond et al., 1963, p. 183). Following the lead of these writers, the terms conditional and unconditional will be used in the present paper in place of the more conventional conditioned and unconditioned.

II. Investigations of Personality Dynamics

A. Studies of Relations between Independently Defined Personality Variables and Parameters of Conditioning Performance

1. Tests of Predictions from Personality Theories of Wide Generality

The only comprehensive theory of personality which has given rise to explicit predictions concerning the differential conditioning performance of groups distinguished on the basis of personality is that of Eysenck. Conditioning is allotted a crucial role in Eysenck's theory. In the development of his theory, Eysenck began with a study of the problems of taxonomy or classification. At this stage the question asked was, "What are the major dimensions of personality with respect to which persons vary?" The answer, proposed on the basis of previous findings and original

research, was that most of the variance in personality functioning can be accounted for in terms of the three orthogonal dimensions of psychoticism, neuroticism, and introversion-extraversion. Psychoticism is defined as a predisposition to develop such symptoms of mental disorder as delusions, hallucinations, mood disturbances, motor retardation, and the like. Neuroticism is identified with emotionality or lability of the autonomic nervous system, which is considered to act as a predisposition to neurotic disorders. The introversion-extraversion dimension is defined in terms of a wide range of behaviors. The behavior of introverts is characterized by a relative lack of sociability, high persistence, high level of aspiration, an emphasis on accuracy rather than speed, reliance on inner standards of conduct, and a stress on moral scruples. Extraverts, on the other hand, are sociable, impulsive, dependent on the social valuations of others, low in level of aspiration, and tough minded in their attitudes. Following Jung, Eysenck proposes that hysteria is the syndrome to be found in the extraverted neurotic, whilst dysthymia (syndrome characterized by anxiety, reactive depression, and/or obsessive-compulsive features) is typically found in the introverted neurotic.

In seeking likely neurophysiological mechanisms of the personality differences between extraverts and introverts, Eysenck formulated his individual difference and typological postulates in terms of excitation-inhibition balance. The essence of these postulates is that individuals in whom excitatory potentials are generated quickly and strongly, and in whom inhibitory potentials are generated slowly and weakly, will tend to be introverted in personality, and to develop dysthymic disorders in the case of neurotic breakdown. Conversely, individuals who generate weak excitatory potentials slowly, and who generate strong inhibitory potentials quickly, tend to be extraverted in personality, and to develop hysterical-psychopathic disorders in case of neurotic breakdown.

A link between excitation-inhibition balance and (a) the personality patterns of introversion *versus* extraversion, and (b) hysterical-psychopathic disturbances *versus* dysthymia, is provided by conditioning.

In brief, a key difference between extraverts and introverts is the degree of socialization which is typical of each. Socialization, or the establishment of social controls over egoistic impulses, is mediated by conditioning. Because of their rapid and strong development of excitation and their weak tendency toward the development of reactive inhibition, introverts condition well, and hence tend to become oversocialized. Conversely, the slow development of weak excitatory potentials, and the rapid and strong development of inhibitory potentials, makes extraverts condition poorly. As a consequence extraverts tend to be undersocialized.

Strong autonomic-emotional lability and reactivity produce excessive

fear reactions to painful stimuli in all persons high on neuroticism, but in the introverted neurotic, the strong capacity for conditioning causes these fear reactions to become attached to a multitude of fortuitous stimuli, thus producing the excessive anxiety reactions of the dysthymic. The dysthymic, then, is characterized by oversocialization and excessive anxiety, whilst the hysteric is characterized by undersocialization and autonomic lability without excessive conditional anxiety.

It is clear that Eysenck's theory postulates a general factor of conditionability or acquisition of CR's. It also leads to the prediction that introverted normals will condition more rapidly than extraverted normals with any reliable conditioning procedure. Furthermore the theory predicts that neurotic introverts (dysthymics) will condition better than unselected normals, and that neurotic extraverts (hysterics and psychopaths) will condition less well than unselected normals. There is no suggestion, however, that neuroticism, anxiety, or emotionality as such will be related to conditioning performance.

The first experiment designed to test these hypotheses was carried out by Franks (1956), who used the eyeblink conditioning method with 20 hysterics, 20 normals, and 20 dysthymics. In accordance with prediction, the groups were ordered dysthymics, normals, and hysterics in terms of rate of conditioning and resistance to extinction. In a later experiment, Franks (1957) used 55 normal undergraduate Ss who had been given the Maudsley Personality Inventory (MPI). This inventory included an extraversion (E) scale. Eyeblink conditioning was again used, and the correlation between conditioning and extraversion was $-.46$ for acquisition, and $-.34$ for extinction.

More recent studies of eyeblink conditioning in relation to scores on the E scale of the MPI have mostly produced relationships in the direction predicted by Eysenck, but few of the correlations have reached statistical significance. Willet (1960) used the Fisher Z transformation to obtain an estimate of the true correlation between eyeblink acquisition score and the E scale. The estimate, which was based on the studies of Franks (1957), Das (1957), O'Connor (1959), and Willet (1960), was $-.188$. Since Willet's publication, further investigations of the relation between the E scale and eyeblink conditioning have included those by Barendregt (1961), Field and Brengelmann (1961), and Sweetbaum (1963). Barendregt reported a correlation of $-.29$ but Sweetbaum obtained a correlation of near zero. Field and Brengelmann found that their introverted Ss conditioned insignificantly faster than their extraverted Ss. It would appear that the results of the later investigations agree with those cited by Willet (1960) in suggesting that the relationship between the E scale and eyeblink conditioning is best represented by a correlation of $-.2$.

Several investigators have studied the relationship between extraversion questionnaire measures and GSR conditioning. Becker (1960) and Becker and Matteson (1961) twice presented a list of 104 words spaced at 5-sec intervals. The word "repeat" occurred 20 times followed by a .1-sec shock on 12 trials. Using two different measures of conditioning, Ss with introverted scores on Guilford's R scale were not differentiated from Ss with low scores on the scale, in either experiment, despite the fact that in the second experiment extreme scoring groups were selected.

Vogel (1960, 1962), however, using a very similar GSR method, obtained highly significant differences in conditioning performance between extraverts and introverts. Lovibond (1963a) also used the GSR in a study of conditioning in 63 university students for whom E-scale scores were available. There were two conditioning sessions. In one session the US was electric shock and in the other the US was provided by pictures of nude females. The correlation between E score and a measure of rate of aversive conditioning was −.05. The correlation for positive reinforcement was −.10.

A change of conditioning technique appears to leave the correlation between the E scale and conditioning performance essentially unchanged, and it is clear that the amount of common variance between the two types of measures is small. Possible reasons for the failure to demonstrate a stronger relationship, other than invalidity of the theory, include unreliability of the conditioning measures, and inadequacy of the E scale as a measure of extraversion. By and large it is clear that the reliability of both eyeblink and GSR conditioning measures is reasonably satisfactory (Becker, 1960; Franks, 1957; Lovibond, 1963b; Welch and Kubis, 1947).

It will be suggested in Section III, D,2 that the validity of GSR conditioning measures can often be questioned, but the arguments advanced there do not apply to the Becker (1960) experiment, nor to Lovibond's (1963a) positive reinforcement measure, both of which failed to produce an appreciable correlation with E scores. Willet (1960) has argued that the E scale is a very poor measure of extraversion, and cites a study of Claridge which showed that careful behavioral ratings produced predicted relationships whereas questionnaire scores did not. Eysenck (Eysenck and Claridge, 1962) has recently recognized that the E scale is not unidimensional, and in fact measures a "behavioral" extraversion factor as well as a "constitutional" extraversion factor, the former being irrelevant to predictions from the original theory. Whilst a purified E scale may improve the position, it is likely that questionnaire studies of personality-conditioning relationships have reached the limit of their fruitfulness. It would now seem advisable to concentrate research effort on the devel-

opment of a set of rating scales along the lines of those developed by Olson (1930) for the study of behavior disorders in children, and to use these scales in conjunction with objective tests of excitation-inhibition balance (Eysenck and Claridge, 1962).

2. *Tests of Predictions from Theories of Limited Aspects of Personality Functioning.*

a. Anxiety and Conditioning. The personality variable which has been most studied in relation to conditioning performance is anxiety. Since the majority of personality theories had their origins in the field of abnormal psychology, where anxiety is a pervasive problem, it is understandable that anxiety has become a central personality concept. As Jensen (1958, p. 313) has put it, "Few, if any, other concepts sustain as much of the superstructure of personality theory."

The early studies of anxiety and conditioning were largely empirical, but since developments of Hull's general behavior theory gave rise to testable predictions concerning the relation between anxiety and conditioning, interest in this area of investigation has increased markedly. Much of the recent work has been carried out by the group at Iowa (Spence, 1954; Spence and Taylor, 1951, 1953; Spence and Farber, 1953; Taylor, 1951, 1956).

The aspects of Hullian theory which are of relevance to the relation between conditioning and anxiety have been stated by Taylor (1956) as follows:

"According to Hull, all habits (H) activated in a given situation, combine multiplicatively with the total effective drive state (D) operating at the moment to form excitatory potential or E $[E = f (H \times D)]$. Total effective drive in the Hullian system is determined by the summation of all extant need states, primary and secondary, irrespective of their relevancy to the type of reinforcement employed. Since response strength is determined in part by E, the implication of varying drive level in any situation in which a single habit is evoked is clear; the higher the drive, the greater the value of E and hence of response strength. Thus in simple, noncompetitional experimental arrangements involving only a single habit tendency, the performance level of high drive Ss should be greater than that for low drive groups."

In a series of experiments designed to test this prediction from Hull's theory, the Iowa group chose classical conditioning of the eyeblink as "a simple, noncompetitional experimental arrangement involving only a single habit tendency." In the search for a drive which could readily be varied in strength, Spence and his associates made the assumption that

anxiety possesses the energizing and reinforcing properties attributed to drive in Hullian theory.

The next step was to provide a means of manipulating the strength of the anxiety drive. There are three principal procedures which experimenters have used for this purpose. The first selects Ss who manifest symptoms of anxiety in different degrees by means of clinical observation or responses to questionnaires. The second induces anxiety experimentally by means of ego threat (e.g., failure in a competitive situation) or threat of painful stimulation (e.g., electric shock). The third method makes use of real life situations considered to be productive of anxiety (imminence of examinations or surgical operations).

The most favored procedure has been the first named, and most experiments have made use of a specially devised questionnaire, the Manifest Anxiety Scale (MAS). The MAS was constructed by asking clinicians to select from a pool of personality questionnaires those items, the endorsement of which would be indicative of manifest anxiety. Although the rationale of this procedure has not been discussed in any detail, it is clear that when the MAS was constructed, it was assumed that high scorers on the scale were in a chronic state of high anxiety drive.

A sufficient number of studies has now been carried out to indicate that MAS scores are positively related to eyeblink conditioning performance (Baron and Connor, 1960; Spence and Farber, 1953; Spence, Farber and Taylor, 1954; Spence and Taylor, 1951, 1953; Taylor, 1951). The strength of this relationship is represented by a correlation of about +.25. In other words, differences in MAS score account for about 6% of the variance in eyeblink conditioning performance.

Bitterman and Holtzman (1952), Gilberstadt and Davenport (1960), and Lacey, Smith, and Green (1955), were unable to demonstrate a significant relationship between MAS scores and GSR conditioning. More recently, however, Becker and Matteson (1961), using extreme scores on the MAS, found that the rate of acquisition of a conditional GSR was substantially greater in anxious Ss than in nonanxious Ss. The relatively low correlations obtained between the MAS and conditioning, as well as the not infrequent failure of scores on the scale to correlate as predicted with complex learning performance, has led to a more critical attitude toward the MAS, and a tendency to use other methods of manipulating the anxiety variable. At the same time the conception of anxiety itself has undergone some change. It is now fairly generally agreed that the MAS does not measure present defensive arousal, proneness to defensive arousal, or chronic, persisting tension, but rather S's tendency to react to anxiety in a way "characteristic of the psychasthenic neurotic" (Jenkins and Lykken, 1957). There seems to have been a

definite shift toward the conception of anxiety as sensitivity to defensive arousal or reactivity to threat. Anxiety thus becomes a disposition rather than a state, but it is still considered to have drive properties when activated by threat or stress. This shift in emphasis in the definition of anxiety has resulted in part from the greater frequency of significant findings in tests of drive theory when threat has been experimentally induced. For example, Beam (1955) made use of the real life stress produced by impending doctoral examinations in a study of GSR conditioning. Under stress, conditioning was significantly faster. Sweetbaum (1963), in a similar type of study, relied on impending surgical operation to induce anxiety. In this case the response conditioned was the eye blink. Subjects under stress conditioned very much more rapidly than did control *Ss* who had made a successful recovery from surgery. Willet (1963) compared the eyeblink conditioning performance of youths taking part in a stressful selection procedure with that of controls who had already been selected. The acquisition of the CR was significantly faster in the experimental group.

Experiments which have made use of clinical ratings of anxiety have also tended to produce more significant results than studies using the MAS. Some of the earlier experiments of this type were not prompted by the theory of anxiety as a drive, but because of their relevance for an evaluation of this theory, they are mentioned here. Welch and Kubis (1947) investigated the effects of pathological anxiety on the rate of GSR conditioning. A group of 51 mixed psychiatric patients was rated for anxiety by psychiatrists. The CS was one of 54 nonsense syllables presented at a rate of one per 6 sec. The US was the sound of a buzzer. The conditioning score was number of trials required for 3 successive GSR's to the CS greater than the GSR's to control syllables. All but 4 of 36 normals scored from 14 to 58, whilst, with one exception, all patients diagnosed as having anxiety had a score of 14 or less. The retest correlation on 36 normals was +.88. Other investigators who have found clinically assessed anxiety to be positively related to GSR conditioning performance are Bitterman and Holtzman (1952), Gilberstadt and Davenport (1960), and Schiff, Dougan, and Welch (1949). No failures to find such a relationship appear to have been reported.

Altogether there is a considerable body of evidence that psychiatric patients suffering from anxiety, and normal individuals subjected to stress, condition more rapidly in experiments using aversive reinforcement than do persons not suffering from anxiety. The theory of Spence and Taylor, however, predicts similar findings from nonaversive conditioning experiments. According to the theory, anxiety as an irrelevant drive should summate with relevant drive to increase performance in

this sort of situation also. In order to test this prediction, Bindra, Paterson, and Strzelecki (1955) undertook a salivary conditioning experiment with high and low scorers on the MAS. The performance of the anxious and nonanxious *S*s was virtually identical. Bindra *et al.* interpreted their results as favorable to the view of anxiety as a response to threatening situations. The findings of Bindra *et al.* have been accepted uncritically into the literature as indicating that anxious and nonanxious *S*s perform equally well in classical conditioning experiments involving no threat (Eysenck, 1957; Franks, 1960; Jenkins and Lykken, 1957; Kimble, 1961; Stewart, Winokur, Stern, Guze, Pfeiffer and Hornung, 1959). The Bindra experiment, however, is open to such serious methodological criticism (see Section III, D, 3) that its results cannot be accepted as evidence for or against any theory. Hence the issue of whether or not anxiety increases performance in nonaversive conditioning must be regarded as still open.

An alternative to the general drive theory of the relation between anxiety and conditioning can be derived from recent work in the field of electrophysiology and Pavlovian theory. As Malmo (1958) has argued, the concept of general drive or *D* may be regarded as identical with the neurophysiological concept of "nonspecific arousal." Since the pioneering work of Moruzzi and Magoun (1949), nonspecific arousal has been seen as the chief function of the brain stem reticular formation. It is now recognized, however, that the reticular formation can no longer be regarded simply as a nonspecific arousal system, the function of which is merely that of an excitatory amplifier. The work of such investigators as Kaada and Johannsen (1960) has made it abundantly clear that the reticular formation selectively activates some areas of the brain whilst it inhibits others. As Diamond *et al.* (1963, p. 228) put it, the work on the reticular formation assumes "its full significance only with the recognition that multiple inhibitory effects, exerted both upwards and downwards, are essential to all behavior."

Microelectrode and intracranial chemical stimulation techniques have demonstrated the existence of "centers," or functional systems associated with unconditional reactions of various types, and have confirmed Pavlov's view of reciprocal inhibitory relationships between these centers. The experiments of Wyrwicka and Dobrzecka (1960) on goats confirmed the hypothesis of Anand and Brobeck (1951) that the lateral hypothalamus contains a "feeding center" and the ventromedial nucleus a "satiation center." Considering this and related experiments, Diamond *et al.* (1963, p. 368) comment that "feeding, as an element of the behavioral repertoire, is embedded in several different behavior pairs or polar opposites. First, we have the contrast between feeding and satiation

as originally postulated. Second, there is a conflict between feeding and fear, . . . third, in the feeding center itself, there is the differentiation between eating and drinking. When an animal chooses not to eat, there may be at least three kinds of motivational influence: satiation, fear, or felt need for liquid. It is an established fact that any one of these inhibits the feeding mechanism." Wolpe (1958), on the other hand, has developed a system of psychotherapy based on animal experiments in which reciprocal inhibition of anxiety responses is effected by the activation of positive responses such as feeding.

Anokhin (1961) has elaborated a conception of the selective action on the cerebral cortex of reticular and other subcortical nuclei in relation to the formation and elicitation of conditional reflexes. Summarizing the results of electrophysiological experiments which made "a comparative evaluation of the activating effect on the cerebral cortex in two biologically opposed states: defensive and alimentary," Anokhin (1961) says, "It may be assumed that all forms of desynchronization of slow cortical activity that arise during the orienting-investigatory, defensive, and alimentary reactions, although outwardly remaining to some extent similar, are actually physiologically entirely specific, ensuring entirely different biological activities . . . the activating influence on the cortex is always of a functionally specific character . . . only the imperfection of the electroencephalographic index fostered the emergence of the ideas of a 'generalized,' 'nonspecific,' and 'diffuse' activating effect of the subcortex on the cerebral cortex."

In the light of these considerations, it would seem reasonable to postulate that individuals in a state of strong anxiety, whether chronic or acute, will manifest a decreased capacity for forming positive reward conditional reflexes. This prediction is based on the assumption that defensive arousal will inhibit the functionally specific arousals considered necessary for the formation of positive reward, or appetitive, conditional reflexes. No clear-cut evidence relevant to this prediction is available. The results of Bindra et al. (1955) are equivocal, and, although Finesinger, Sutherland, and McGuire (1942) have shown that neurotic Ss are able to form salivary conditional reflexes, no control data were obtained. Lovibond (1963a) compared the positive reward GSR conditioning performance of high and low scorers on Eysenck's Neuroticism (N) scale, which had previously been shown to correlate highly with the MAS (Eysenck, 1957). There was an insignificant tendency for Ss with low N scores to perform at a higher level than Ss with high N scores.

With the redefinition of anxiety in terms of emotional sensitivity to threatening situations, it is now clear that experiments using questionnaire scores as anxiety criteria are of little relevance in the study of

the relationship between anxiety and appetitive conditioning. What is required is a study of the effects of real life stress, such as impending surgical operation (Sweetbaum, 1963), on appetitive conditioning performance using either the GSR technique of Lovibond, or a methodologically acceptable technique of salivary conditioning (see Section III, D, 3).

We have seen that the two theories which have generated most of the research into personality conditioning relations emphasize different personality variables as being related to differences in conditionability. For Eysenck, the personality dimension which is related to conditionability is introversion-extraversion whilst for Spence and Taylor it is anxiety. We have seen also that both variables in fact appear to some degree to be related to aversive conditioning performance. Hence the question of the relation between these two variables naturally arises. It will be remembered that Eysenck's theory of personality postulates three orthogonal dimensions, introversion-extraversion, neuroticism, and psychoticism. Neuroticism has been defined tentatively in terms of emotionality, or autonomic nervous system lability, and since Eysenck has in mind defensive emotionality and autonomic lability, it would appear that neuroticism and anxiety may reasonably be equated. The essence of Eysenck's thinking on the subject appears to be conveyed if we think of anxiety-neuroticism in terms of "defense arousability," i.e., sensitivity to arousal by aversive stimulation and threatening situations.

Eysenck has postulated that introverts generate excitatory potentials rapidly and strongly. His findings that, in contrast to extraverts, introverts have high levels of aspiration, persistence, perceptual rigidity, and sedation threshold, support this assumption, although Eysenck has emphasized the differences in inhibitory potential between introverts and extraverts rather than differences in excitatory potential.

A reconcilation between the views of Eysenck and Spence and the accumulated evidence may be effected by acceptance of the following propositions.

1. The level of cortical excitation, which determines the speed of formation of conditional linkages, is dependent on
    (a) the degree of subcortical arousal and
    (b) the general excitability of the cortex, i.e., the capacity of the cortex for developing excitation from a given level of subcortical arousal.

2. Individuals differ with respect to three dimensions of nervous activity which are at least partially independent:
    (a) general excitability of the cortex,

(b) defensive arousability, and

(c) appetitive arousability.

3. The personality dimension which is related to general excitability of the cortex, and hence to a general factor of conditionability, is introversion-extraversion, and the relationship is such that introverts will manifest higher levels of general cortical excitability, and conditionability, than will extraverts.

4. The personality variable of anxiety is related to level of defensive arousability, and a group factor of aversive conditionability.

5. The relationship between different arousal systems is a mutually inhibitory one, so that, e.g., a high level of defensive arousal at a particular time will reduce the capacity for appetitive arousal, and hence appetitive conditioning performance.

6. The relationship of reciprocal inhibition between different types of arousal will be to some degree asymmetrical. Because of the generally greater strength of defensive arousal, the inhibitory effects of this type of arousal are likely to be more readily observable than the inhibitory effects of nondefensive arousal.

Table I sets out the predicted differences in the aversive and appetitive conditioning performance of anxious introverts, anxious extraverts,

TABLE I

CONDITIONING PERFORMANCE OF VARIOUS GROUPS EXPECTED ON THE HYPOTHESIS OF A RELATION OF RECIPROCAL INHIBITION BETWEEN DIFFERENT AROUSAL SYSTEMS

	Anxious (stress situation)				Nonanxious (nonstress situation)			
	Aversive conditioning		Appetitive conditioning		Aversive conditioning		Appetitive conditioning	
	General	Specific	General	Specific	General	Specific	General	Specific
Introvert	+	+	+	—	+	0	+	0
Extravert	—	+	—	—	—	0	—	0

nonanxious introverts, and nonanxious extraverts. It is important to note that "anxious" refers here to a presently high level of defensive arousal, whether this is achieved through a high degree of defensive arousability and moderate stress, or a moderate degree of defensive arousability and a high level of stress.

Considering the anxiety dimension alone, the predictions are quite clear cut and readily tested. It is predicted that an unselected group under real life stress will condition better than nonstress controls with aversive reinforcement, but less well with appetitive reinforcement. We have seen that the evidence of Beam (1955), Sweetbaum (1963), and

Willet (1963) is favorable to the prediction relating to aversive conditioning, but evidence relevant to appetitive conditioning is entirely lacking.

Considering the introversion-extraversion dimension separately, it is predicted that introverts will condition more readily than extraverts with both aversive and appetitive reinforcement. The only unambiguous studies bearing on these predictions are those which have used questionnaires to select Ss. In all cases except one (Lovibond, 1963a), aversive conditioning procedures have been used, and the evidence, on balance, favors the prediction of superior performance by introverts. In Lovibond's (1963a) appetitive GSR conditioning experiment, there was a tendency for introverts to perform at a higher level than extraverts. This performance difference, although in the required direction was not significant.

It was suggested in Section II, A, 1 that the selection of extraverts and introverts by means of personality questionnaires alone is an unsatisfactory procedure. Efforts to obtain behavioral ratings which emphasize the temperamental disposition factor of extraversion might well prove fruitful, particularly if combined with objective indicators such as level of aspiration, persistence, and so forth. Extraverts and introverts selected by these means could be equated on the anxiety dimension by means of questionnaires and measures of autonomic reaction to stress.

Turning now to the predictions derived from the interaction of the two dimensions, it will be seen that, in Table I, pluses, zeroes, and minuses are allotted under the various headings. A plus signifies better-than-average conditioning expected, a zero indicates average conditioning, and a minus below-average conditioning. The heading "general" refers to the expected influence of the assumed general factor, and the heading "specific" refers to the expected influence of defensive arousal.

Assuming equal weight for each positive and negative tendency, the algebraic sums of general and group factor influences give the expected order of the groups in terms of each type of conditioning performance. The decreasing order of predicted aversive conditioning performance for the groups is anxious introvert, nonanxious introvert, anxious extravert, nonanxious extravert. The corresponding order for appetitive conditioning is nonanxious introvert, anxious introvert, nonanxious extravert, anxious extravert.

b. Conceptual Thinking Disturbances and Conditioning. From studies of the responses of schizophrenics and normals to various tests of concept formation, Lovibond (1954) concluded that the characteristic feature of schizophrenic thought disorder is a failure to inhibit linkages or associations which the context makes irrelevant. According to this hypothesis, in normal thinking, the context, including the experimenter's

instructions, the nature of the test materials, or the theme of a discussion, permits the activation of only a limited number of "relevant" linkages in the S's conceptual structures. The majority of the manifold linkages that are potentially available are selectively inhibited by a process akin to the active, differential inhibition manifest in conditioning experiments. On the basis of these assumptions, a method was developed to permit analysis of responses to the Object Sorting Test in terms of the in-essentiality of the definitions offered (Lovibond, 1954, 1964a). Further work with this test indicated that something like 20% of normals obtained high, or schizophreniclike, scores. It was concluded that certain aspects of schizophrenic-type thinking could conveniently be studied in these high-scoring normals whose thinking was free from the complications of mental disorder.

In an extension of the original theorizing, it was hypothesized that the differential inhibition of irrelevant linkages in normal thinking takes place through a process of negative induction, which necessitates a relatively strong process of excitation. On the basis of this analysis, a number of predictions was made concerning the differential conditioning performance of normals with high and low scores on the Object Sorting Test. In particular, it was predicted that high scorers would show lower reactivity (including weaker initial orienting reactions, and a slower rate of conditioning) and a lower capacity for differentiation. With use of positive, or appetitive, reinforcement of the GSR, the main predictions were confirmed. When the GSR was reinforced by electric shock in another session, however, the differences between the two groups largely disappeared. It was concluded that the stronger activation from the aversive US was sufficient to mask the differences in conditioning performance between the two groups.

This study raises the question of the relationship between the personality dimension tapped by the Sorting Test, and other dimensions known to be related to conditioning performance. The Sorting Test dimension would seem to be closely akin to the "cortical excitability" dimension assumed by Eysenck to be one of the properties of nervous activity which differentiates introverts and extraverts. The finding of stronger differential inhibition in Ss with strong excitatory processes introduces a contradiction, however, since Eysenck's work suggests a negative correlation between strength of excitation and strength of inhibition. Nevertheless, it is possible that this contradiction is more apparent than real. Eysenck has been concerned primarily with the "work decrement" type of inhibition, which may be not at all the same process as the active inhibitory process assumed by Pavlov to underlie differentiation. In this connection it is of interest to note that, in differential delayed condition-

ing, when inhibition with reinforcement occurs, the response to the differential stimulus usually shows little further decrease, and may actually increase. This suggests the possibility that response decrement inhibition may be inimical to differential inhibition. No consistent correlations have been obtained between Sorting Test scores and scores on the MPI, but in view of the criticisms of this and related questionnaires, it would be unwise to conclude that the tests are measuring different dimensions.

The correlation of Sorting Test scores and measures of extraversion based on ratings and objective tests might well produce a different result.

3. Empirical Studies of Personality and Conditioning

In a number of purely empirical studies, comparisons have been made of the conditioning performance of groups known to differ in terms of certain personality characteristics. Usually the groups have been distinguished on the basis of psychiatric diagnosis or symptomatology.

The earliest reported studies of conditioning in psychiatric patients were carried out by Russian investigators, including Ivanov-Smolensky (1925), Guk (1934), and Landkof (1938). Usually very small numbers of subjects were used, but the invariable finding was that, with psychotic subjects, conditional reflexes were difficult to establish, and, when finally established, the reflexes were quite unstable. The reflexes conditioned in these studies included defensive, motor, and avoidance reactions.

Bender and Schilder, (1930) in agreement with the Russian findings, reported difficulty in establishing conditional avoidance reactions in a small group of schizophrenics. Mays (1934), however, found that conditional GSR's were more resistant to extinction in a group of 20 schizophrenics than in 16 normal controls. Shipley (1934) studied the acquisition and resistance to extinction of conditional GSR's in psychiatric patients. The Ss were 17 schizophrenics, 10 manic depressives, 9 mixed neurotics, and 6 normals. The schizophrenics were found to condition at the fastest rate, with normals next, followed by the manic depressives and neurotics. The neurotic Ss were the most resistant to extinction, with the other groups undifferentiated on this measure.

Pfaffman and Schlosberg (1936) used a knee jerk conditioning procedure with schizophrenics, manic depressives, and normals. The schizophrenics tended to be the least responsive group, but the over-all level of conditioning was too low to permit definite conclusions.

Peters and Murphree (1954) investigated differential GSR conditioning to shock in schizophrenics and normals. The difference in response

to the conditional stimulus and the differential stimulus after training was markedly less in the schizophrenics.

Taylor and Spence (1954) compared the rate of eyeblink conditioning of 74 Ss suffering from neurosis or personality disorder, and a group of 42 psychotics, including 31 schizophrenics. The difference in performance favoring the psychotics just failed to reach the 5% level of significance.

In a large-scale study of GSR conditioning, Howe (1958) used 60 hospitalized Ss with anxiety states, 60 hospitalized chronic schizophrenics, and 60 normal Ss. Taking magnitude of response during extinction as the measure of strength of conditioning, Ss with anxiety states showed stronger conditioning than schizophrenics or normals. The difference between the latter groups was not significant.

In a further study of GSR conditioning in psychiatric patients, Stewart et al. (1959) used 70 Ss, comprising 27 manic depressives in the depressed phase, 18 schizophrenics, 15 Ss with personality disorders, and 10 with anxiety neurosis. The US was shock, and a delayed conditioning procedure was used with a 4-sec CS–US interval. GSR's were classed as orienting responses, responses to shock, or anticipatory CR's. It was found that Ss with personality disorders conditioned more rapidly than either schizophrenic or manic depressive Ss, and that Ss with anxiety neuroses tended to condition more rapidly than those with schizophrenia. The resistance to extinction of personality-disordered Ss was significantly greater than that of the schizophrenics, but no other group differences were significant. The frequency of orienting responses to the CS during training was significantly greater in the anxiety neurotics than in either of the psychotic groups.

Another finding of interest was that the rate of adaptation to the CS before pairing with the US began was significantly slower in the anxiety neurotics than in either schizophrenics or manic depressives.

The contradictions in the reported findings which have been reviewed, and the marked differences in the procedures used in the various studies, make interpretation somewhat hazardous. Nevertheless, the evidence points to a number of conclusions which perhaps need not be too tentative. In the first place those studies which have used neurotic Ss suffering from symptoms of anxiety have added further weight to the already established fact that anxious persons tend to acquire aversive CR's more rapidly, and to extinguish them more slowly than other groups. Apart from one or two inexplicable discrepancies, it appears that schizophrenics tend to be less generally responsive to either mild or strong stimuli, and also to condition rather less readily than normals. The conclusion concerning the lack of responsiveness of schizophrenics is in

accord with a considerable amount of evidence of a nonconditioning nature. There is quite a strong suggestion that schizophrenics tend to be deficient in "arousal" or "drive," and that this manifests itself in a somewhat retarded rate of conditioning (Lovibond, 1963a). In view of Meehl's (1962) suggestion that schizophrenics may show a deficiency of functioning of the positive reinforcement centers, it is interesting to speculate whether schizophrenics would not show a more marked deficiency in positive reward conditioning than in aversive conditioning. Lovibond's (1963a) findings with normals who gave schizophreniclike responses to the Object Sorting Test (see Section II, A, 2, b) suggest that this may well be the case.

The finding that psychiatric patients, other than schizophrenics and anxiety neurotics, usually condition less well than normals, possibly reflects a somewhat lowered cortical efficiency which is common to all patients, but which sometimes does not manifest itself if the patient has heightened drive relevant to the response being conditioned.

B. Studies in the Definition of Personality Variables through Conditioning

Very recently, some noteworthy advances appear to have been made by Soviet investigators in the development of Pavlov's conceptions of types of higher nervous activity. Of particular interest is the work of Teplov and his associates (Nebylitsyn, 1961, 1963; Nebylitsyn, Rozhdest-venskaya and Teplov, 1960; Rozhdestvenskaia, Nebylitsyn, Borisova and Ermolaeva-Tomina, 1963; Teplov, 1961).

The essential difference between the approaches of Soviet and Western workers to problems of personality and conditioning can be illustrated by a comparison with the work of Eysenck (see Section II, A, 1). Eysenck's initial investigations were made at the level of personality functioning. The first task was to determine the major dimensions of personality. The next step was to seek, at the neurophysiological level, for likely mechanisms of the differences observed at the personality level. Predictions derived from the theory of neurophysiological mechanisms were then tested by the procedures of experimental psychology, including conditioning.

The method of the Soviet workers is quite different. For them the starting point is at the level of hypothetical neurophysiological mecha-nisms suggested by Pavlov's theory of brain functioning and general behavior. Laboratory procedures, with an emphasis on conditioning, are used in an attempt to define the major dimensions with respect to which people vary at this level. The next step in the program is to move to the level of personality, and there to seek correlates of differences at

the neurophysiological level. Teplov's position with respect to the relation between personality theory and general behavior theory has been very clearly stated: "The problems of the psychology of personality are first and foremost problems of general psychology" (Teplov, 1961, p. 21).

Teplov believes that, before proceeding to questions of typology, it is necessary to carry out a much more thorough investigation than hitherto of the basic properties of the nervous system—the dimensions of higher nervous activity. The starting point of a series of investigations was Pavlov's conception of the fundamental properties of nervous activity: strength, balance, and mobility. Instead of taking these properties as given, Teplov raised the question: In the higher nervous activity of human beings, is each of these properties a single, unidimensional property, or a group of similar properties?

Strength was the first property investigated. For this purpose it was necessary to develop a number of laboratory test methods, including conditioning procedures. Most of these techniques are based on the assumption that the physiological essence of the strength property is maximum nerve cell capacity. The "method of reinforced extinction of the photochemical reflex, with and without caffein" is fundamental. The essence of this method (Rozhdestvenskaia *et al.*, 1963) is as follows. A conditional reflex reduction in absolute visual sensitivity of the dark-adapted eye is elaborated to a visual or auditory CS, using a very strong light as US. After consolidation of the CR, its magnitude is measured in one "experiment." (In Soviet literature it is often necessary, as in this case, to read for "experiment" the word "trial.") The measure of CR magnitude is percentage diminution of sensitivity to the CS relative to the background level of sensitivity. Ten reinforcements or CS–US combinations are then given with 2-min intervals, followed by a further measure of CR magnitude. In another session the procedure is repeated after the administration of 0.2 gm caffein. The measure of strength derived from this method is the percentage change in magnitude of the CR from the first to the second test trial, i.e., following "reinforced extinction." A diminution of the CR on retest is an indicator of weakness, or in other words the indicator of strength is the amount of resistance to "inhibition of reinforcement." The administration of caffein is designed to increase the excitability of the nerve cells, and hence to reduce their capacity to withstand prolonged stimulation.

Other techniques are based on the assumption that strength will manifest itself in concentration of excitation and positive induction effects under circumstances where weakness will lead to irradiation, and diminished, rather than increased, response.

The research of Rozhdestvenskaia *et al.* (1963) may be taken as

illustrative of work in this area. In this study, 17 techniques were employed with each of 40 Ss to determine nervous system strength. Each of these methods had been individually studied and the theoretical validation (construct validity) of each was generally accepted. In addition, four measures of magnitude and rate of extinction of vascular orientation responses were included. The 17 strength measures included four measures of reinforced extinction of the photochemical CR (visual and auditory CS, with and without caffein) and four variants of the "induction method" (measures of effects of a punctuate extraneous stimulus on sensitivity to a second punctuate stimulus). Four other measures were derived from ergographic recordings under conditions of diminished nerve cell capacity, achieved by the administration of caffein, and muscular fatigue.

Centroid factor analysis of rho rank order correlations revealed a clear strength factor. Thirteen of the 17 strength measures showed high loadings with respect to this factor. A second factor was tentatively identified as "nervous system stability." The only measures showing high loadings with respect to Factor II were the four vascular orientation reflex indices. A third factor, loading only "photochemical reflex to acoustic stimuli with and without caffein," was interpreted as a special intensity effect of the auditory CS.

Nebylitsyn (1961) has published an account of a similar factor analytic study in which the principal strength measure was inhibition with reinforcement of the conditioned alpha blocking response of the EEG.

The results of studies such as these have led Teplov and his associates to conclude that strength is a unidimensional property of the nervous system. The position is not nearly as clear with respect to the properties of mobility and balance. According to Teplov (1961, p. 36), theoretical considerations and factual evidence suggest that the property of mobility "includes all the time characteristics of the work of the nervous system, and all aspects of this work in which classification by speed can be applied." Possible indices of mobility suggested by Teplov include (a) the speed of development and arrest of nervous processes, and (b) the speed of replacement of excitation by inhibition, and of inhibition by excitation. According to Nebylitsyn (1962), indices of mobility have not been found to correlate highly, and the conviction is growing that it is not a unidimensional property.

Teplov (1961) takes the view that balance is a secondary, derivative property of the nervous system, "for the simple reason that here it is a question of balance in strength or mobility . . . . It is now hardly a matter of doubt that there can be disequilibrium in the mobility of the nervous

processes" (p. 35). Nevertheless, Nebylitsyn (1963) has carried out an investigation based on the hypothesis that balance is a unitary property. In this study, seven indicators of balance were used with 22 student subjects. On the assumption that "one must give preference to the 'conditional reflex (involuntary)' indicators of the relationship among nervous processes," two reference indicators were defined. These were (a) the rate of extinction of the conditional-reflex depression of the alpha rhythm elaborated in the course of "activating" reinforcement (see Section III, D, 4) and (b) the rate of elaboration of differentiation to an acoustic stimulus in alpha depression conditioning. The other indicators were a strength of CR measure, a measure of response to the "unconditional" light stimulus, and measures of strength and rate of extinction of the initial orienting response (alpha depression) to the acoustic CS.

The results revealed that all of the intercorrelations between measures, except two, were significant at the 5% level or better. It was concluded that the results supported the hypothesis that a common property of the nervous system, that of equilibrium, underlay the various response measures. Dominance of excitation over inhibition "causes greater resistance to extinction of alpha rhythm depression, orientation, and conditional reactions, retardation of differentiation, and longer depression of the alpha rhythm upon application of orienting, conditional and reinforcing stimuli" (Nebylitsyn, 1963, p. 27). When excitation and inhibition are in equilibrium, or inverse relationship, the opposite picture is sometimes observed.

Presumably, as indicators of equilibrium become more firmly based, it will be possible to begin to answer the question of the extent to which strength and equilibrium, or balance, are independent dimensions of nervous activity. Certainly the orienting reflex activity measures used by Rozhdestvenskaia et al. (1963), which were similar to those of Nebylitsyn (1963), were virtually independent of the main strength measures.

The mobility concept is likely to continue to be recalcitrant as there seems to be considerable overlap between Teplov's suggested mobility indicators, which were referred to earlier, and the present measures of equilibrium.

A further question may be raised at this point. What is the relationship between "strength," measured in terms of capacity to withstand intense and prolonged stimulation, and what would ordinarily be thought of as "strength of excitation" and which would be measured in terms of amplitude of response? Since it has been shown in a number of studies that strength and sensitivity (in terms of thresholds) are inversely related, it might be expected that strength in these two senses would be inversely related also. In the study of Rozhdestvenskaia et al. (1963) there were two

amplitude measures: the mean levels of the vascular orienting reactions to the auditory stimulus and to the visual stimulus. Neither of these measures correlated at all highly with any of the measures of capacity to withstand adverse conditions of stimulation. However, the auditory and visual vascular amplitude measures correlated only .29 with each other. This low correlation suggests low reliability in one or both of these measures.

It is not surprising, nevertheless, that the two vascular magnitude measures correlated reasonably well with their respective extinction measures, since the two types of measures were not really independent. That is to say, if, as in this case, the measure of extinction is the level of response after a given number of applications of the stimulus, or the number of stimulus presentations necessary to reach zero response, the extinction measures of Ss with initially high levels of response will be relatively low, unless they show a very rapid decrease in response with continued stimulus presentations. Usually this will not occur to any marked degree, and the result will be an inverse relationship between initial response level and "rate of extinction." In Nebylitsyn's case, such a result makes possible an interpretation in terms of predominance of excitation. If, however, the slope of the extinction curve is used as a measure of rate of extinction, a positive relationship may be obtained between initial magnitude of the orienting reaction and its rate of extinction (Lovibond, 1963a). Clearly in this case one can no longer talk in terms of predominance of excitation or inhibition. It should be noted that Nebylitsyn is not alone in using the type of extinction measure he favors, since this measure is conventional in the Western literature.

Despite contradictions and queries such as those mentioned above, one cannot but be impressed by the work of Teplov and his associates. The character of this work has been stated modestly by Teplov: "We are striving to obtain such information about the individual differences in people as will form some system. We are attempting to put forward hypotheses which can be verified experimentally. We have chosen the path from the physiology of higher nervous activity to psychology. We do not consider that this is the only possible path, but we are firmly convinced that it is one of the possible paths." To this might be added the observation that the path being followed is potentially a very rewarding one, and one hopes that as more details of the work become available, Western psychologists will be induced to join in its exploration.

III. Problems in the Investigation of Personality Dynamics by Conditioning Procedures

A. Nonassociative Factors and the Measurement of Conditioning

It has long been known that not all changes in the response measures which occur during conditioning represent "true conditioning," or "experimental extinction," in the sense of being a direct function of CS–US pairing and unpairing. It is customary to describe as "nonassociative factors" any influences on response measures which do not result from the association (or subsequent dissociation) of CS and US. Nonassociative factors giving rise to response increment have been referred to as pseudoconditioning and reflex sensitization. Nonassociative factors producing response decrement have been termed habituation and adaptation. All of these factors, but particularly those leading to response increment, tend to confound measures of CR acquisition.

It has become conventional to refer to the control procedure of presenting the US only during the conditioning phase as pseudoconditioning control, whereas presentation of the CS and US separately is termed sensitization control. It is unlikely, however, that different mechanisms are responsible for pseudoconditioning and sensitization, and it is convenient to use the term pseudoconditioning to refer to nonassociative response increment in general. The precise nature of pseudoconditioning effects is still a matter of controversy (Martin, 1962).

One view is that pseudoconditioning results from a reduction of peripheral response thresholds (Davis, Buchwald, and Frankmann, 1955). Wickens and Wickens (1942) have argued that pseudoconditioning is a process essentially the same as that of conditioning proper. These authors emphasize the associative elements which must necessarily be present in any pseudoconditioning experiment, e.g., the similarity of the situation in which the CS and US are presented. Recent electrophysiological studies (Voronin and Sokolov, 1960) suggest that nonassociative response increment is often a manifestation of the return of the generalized orienting reflex when reinforcement is introduced. Autonomic responses, such as GSR, heart rate change, and vasomotor responses, tend to occur as part of a generalized orienting response to any sudden stimulus change. Consequently, when the CS is presented alone during habituation or adaptation trials, the GSR, for example, occurs as an "unconditional" reaction. Repeated presentation of the CS usually results in a marked diminution in the amplitude of the response to the CS. Voronin and Sokolov (1960) have produced evidence that this adaptation of the orienting reflex results from the development of a process of inhibition, and not merely

from a decrease in arousal. Presentation of any new stimulus after adaptation disinhibits the orienting reflex, and produces a marked increase in response to the CS. If, however, the new stimulus is an aversive US, such as shock, the increased response to the CS is of greater amplitude, and is more prolonged, indicating an increase in arousal as well as disinhibition of the orienting reflex. As conditioning proceeds with further CS–US pairings, the generalized orienting reflex gives way to more specific arousal (Voronin and Sokolov, 1960). If the response being conditioned is an indicator of generalized arousal, there will be a resultant tendency toward response decrement during this stage. Presentation of isolated test stimuli, in a context of repeated CS–US pairings, will again tend to disinhibit the generalized orienting response, as will the change from 100% reinforcement to an extinction series.

It has often been assumed that pseudoconditioning, or nonassociative response increment, is limited to conditioning involving aversive stimulation (Kimble, 1961). Recent work has suggested that this type of response increment may also occur with "neutral" reinforcement, as in conditioning of the alpha blocking response, and with positive reinforcement, as in salivary conditioning (see Section III, D).

The investigation of personality dynamics by conditioning methods would benefit greatly from detailed investigations of the mechanism of nonassociative factors in conditioning. It is unlikely, however, that the operation of these factors can be elucidated fully within the confines of the typical "acute," or very short-term conditioning experiment. Meanwhile, on the reasonable assumption that pseudoconditioning follows laws which differ from those of conditioning proper, investigators of personality and conditioning must seek ways of eliminating nonassociative influences from their response measures. Possible ways of achieving this aim are considered in the discussion of specific conditioning procedures in Section III, D, where it is suggested that the methods of delayed and differential conditioning offer partial solutions to the problem of obtaining uncontaminated conditioning measures.

B. The Influence of Central "Set" Factors

Although classical conditioning is concerned with the modification of involuntary responses, performance during conditioning experiments may be markedly influenced by complex central factors which are usually described in terms of "set," "expectancy," or "attitude." The influence of such factors is perhaps most striking when the response to be conditioned may be performed both voluntarily and involuntarily. In eyeblink conditioning, for example, sets induced by instructions to "let your reactions take care of themselves" or "be sure not to wink before you

feel a puff," can produce very clear-cut facilitatory or inhibitory effects.

In an experiment on the effect of instructions on differential eye-blink conditioning, Hilgard and Humphreys (1938) used both supporting and antagonistic instructions. They found that responses increased in frequency and amplitude and decreased in latency—when they were ordered—from voluntary restraint, through uninstructed, to voluntary facilitation. It should be emphasized that these effects occur in relation to involuntary responses, i.e., responses outside the range of latencies appropriate to voluntary self-induced responses. Furthermore, voluntary restraint does not eliminate conditional responding.

When eyeblink conditioning is used in personality study, it is usual to give facilitatory instructions, but it is not a simple matter to determine the extent to which the instructional set is counteracted by self-induced sets. Razran (1936) suggests that, in salivary conditioning, attitudinal factors are responsible for greatly increased variability of response. Stability of performance is enhanced if Ss have no knowledge of conditioning, or if they can be kept unaware of the nature of the experiment by engaging them in some extraneous task.

Al-Issa (1963) has assumed that "if the inconsistency and variability of adult human conditioning is due to the influence of the subject's verbal and attitudinal activities, then by the mere absorption of the subject in performing some task, these interferences may be ruled out and conditioning may become regular and reliable. These attitudinal interferences are reduced because the absorbing activity takes the subject's attention from the CS and the US and thus he will have less voluntary control over the activities related to conditioning." Accordingly, in an eyeblink conditioning experiment conducted by Al-Issa (1963), one group had the task presented to them as a reaction-time study. Subjects in this group were required to respond to the CS as rapidly as possible by pressing a Morse key. Key pressing was found to result in significantly higher levels of responding.

In an early study of the effect of instructions on the GSR, Cook and Harris (1937) gave their Ss a series of adaptation trials with the stimulus to be conditioned (light). No mention was made of the fact that an electric shock was to be used in the experiment. One group of Ss was then told that electric shock would follow the light on subsequent trials. A further test with the CS alone showed a greatly increased level of response in this group. The subsequent magnitude of response did not increase further with 30 CS–US pairings. Cook and Harris further investigated the effect of instructions on extinction. They found that when Ss were told that the light would no longer be followed by shock, the GSR to the light was very rapidly eliminated in comparison with controls.

The effect of *S*'s expectations on his reactivity is illustrated in a study by Lovibond (1963a). Negative reinforcement of the GSR (shock) was used in one session, and positive reinforcement (nude pictures) was used in another. The test sessions were approximately 1 week apart, and order of sessions was counterbalanced. Before the beginning of each session *S*s knew whether shock was to be used or not. The response to the first adaptation trial of the CS was approximately 50% higher in the shock session than in the nonshock sessions.

In a recent study of the effect of instructions on GSR conditioning, Silverman (1960) showed that the CS–US interval interacts with the effects of instructions. Wickens, Allen, and Hill (1963) found that instructions to the effect that no more shocks are to be presented do not immediately eliminate GSR's to the CS, but rather result in a faster rate of extinction.

It is clear that the course of conditioning, differentiation, and extinction can be influenced by central processes arising partly out of contextual stimuli, including instructions, and partly out of the varied attitudes and expectations which *S*s bring to the laboratory. It seems probable that most of these influences are mediated by language, and hence it is likely that their laws of operation differ from those governing conditioning of simple responses. If this is the case, it follows that these factors must serve to increase the error variance in conditioning measures. Further, since these central factors undoubtedly play a more important role in some conditioning procedures than in others, the instrusion of these influences will tend to vitiate cross-technique comparisons of conditioning. Nevertheless, it is easy to exaggerate the influence of complex central factors. Most of the dramatic effects obtained with GSR conditioning (such as rapid extinction when no shock instructions are given, and increased response to the CS when *S*s are merely told of impending shock after the CS) act on nonassociative components in the response measures. As we gain more understanding and control of non-associative factors in conditioning, our understanding and control of central set factors should likewise increase.

C. GENERAL FACTOR OF CONDITIONABILITY

A problem which is frequently raised in relation to studies of conditioning and personality is the extent to which a general factor of conditionability, or ease of conditional reflex acquisition, underlies performance in the various types of conditioning situations commonly employed. It is argued that theories of personality, such as that of Eysenck, assume the existence of a strong general factor, but such a factor has not been shown to exist. In fact, measures of conditioning have seldom been found to correlate at all highly. At first sight it would appear to be

a simple matter to undertake a factor analytic study of conditioning, with a variety of techniques, to decide the issue of the existence and strength of a general factor. There are, however, many diffculties in the way of a definitive general factor study, and some of the more important ones will shortly be taken up. Before turning to these problems, the possible existence of group and specific factors of conditionability will be considered.

1. Functionally Specific Arousal Systems and the Possibility of Group and Specific Factors

In a preceding section (II, A, 2, a), the probable existence of functionally specific arousal systems, in addition to a general arousal system, has been discussed. Whilst a degree of intraindividual consistency in the reactivity of these systems is to be expected, such consistency is likely to be far from perfect. To the exent that functionally specific arousal systems are of importance in the establishment of conditional linkages, intraindividual variability in the reactivity of these systems will reduce the strength of any general factor of conditionability.

It seems then that it is reasonable to expect a general factor of conditionability, and, in addition, specific and/or group factors. It is possible, if not probable, that there exists group factors associated with the broad defense-arousal and positive reward systems, in other words, factors related to aversive and appetitive conditioning. It is clear, however, that differentiations exist within these broad systems (attention has already been drawn to the division within the feeding center). The work of Olds (1956) is of interest here. Using the technique of self-stimulation through implanted electrodes, Olds mapped out areas centering on the hypothalamus within which mild electrical stimulation has either positive or negative reward value. In later experiments, Olds (1958) investigated the possibility of differentiation within the positive reward areas. Response rates with given electrode placements were determined under conditions of high and low hunger and sex drive. The results pointed to the existence of a localized hunger-reward system differentiated from a local sexual reward system with a reciprocally inhibitory relationship between the two. It is possible that each arousal-reward system will be found to be associated with a specific factor of conditionability.

There are many reasons that the low intercorrelations which have been found between different types of conditioning measures do not constitute decisive evidence against the existence of an important general factor of conditionability (Eysenck, 1962). Among the many influences

which, in addition to specific conditionability factors, will tend to attenuate observed correlations, may be listed the following.

a. Variability in Peripheral Response Mechanism Sensitivity. Since the development of a central conditional linkage can be measured only through the medium of some peripheral response mechanism, such as blinking of the eyes or activity of the sweat glands, it follows that any intraindividual variability in the sensitivity of the peripheral mechanisms, which is not highly correlated with variability in conditionability, will reduce the correlations between conditioning measures.

b. Nonassociative Contamination of Conditioning Measures. To the extent that nonassociative factors not highly correlated with conditionability enter differentially into various types of conditioning measures, the observed relationship between the measures will be spuriously low. For example, different degrees of such confounding in negatively reinforced and positively reinforced GSR's almost certainly reduced the correlation obtained by Lovibond (1963b) between these two types of conditioning.

c. Central Set Influences in Conditioning Measures. In the discussion of central "set" and "attitude" factors in relation to conditioning, it was pointed out that different types of conditioning are susceptible to these influences to different degrees. Since set and attitude are unlikely to covary consistently with conditionability, these factors represent another possible source of attentuation of correlations between different types of conditioning measures.

d. Procedural Variations. Among the many possible procedural dimensions with respect to which conditioning experiments can and do vary, the following appear to be some of the most important: (i) Strength of US; (ii) CS–US interval; (iii) length of intertrial interval; (iv) nature of the intertrial interval (silent, filled, or partially filled (see Section III, D, 2); (v) percentage of reinforcement.

The influence of variations in the above aspects of procedure are far from being fully determined for any one type of conditioning. Sufficient is known of their operation, however, to suggest that, at least to some degree, optimal conditions are likely to vary with the type of conditioning. For example, intertrial intervals of around 20 sec, which appear not to be detrimental to eyeblink conditioning, are too short for optimal GSR conditioning (Prokasy, Hall, and Fawcett, 1962). Furthermore, the interactions between procedural variations are, as might be expected, far from simple. For example, although a CS–US interval of about 450 msec has been shown to produce maximum performance in the conditioning of finger withdrawal (Fitzwater and Thrush, 1956; Spooner and Kellog, 1947), eyeblink conditioning (Kimble, 1947; Kimble, Mann

and Dufort, 1955; McAllister, 1953), and GSR conditioning (Moeller, 1954; White and Schlosberg, 1952), a recent study by Kimmel and Penny-packer (1963) has shown that differential conditioning of the GSR increases as a function of CS–US interval up to the maximum value studied (2 sec). Caution is in order, however, in interpreting GSR studies such as those of White and Schlosberg (1952), and Moeller (1954), because of the probability that conditional and orienting GSR's were confounded in the response measures.

Pavlov (1927) apparently found no significant differences in ease of salivary conditioning between CS–US intervals up to 5 sec, as CR's formed with these intervals were all termed "simultaneous" CR's. No relevant data on human salivary conditioning is avaliable, and the issue of whether or not the optimal CS–US interval is the same for all types of conditioning remains open.

For the purposes of investigating a general factor of conditionability, it would appear desirable to seek optimal procedures for each technique, rather than to attempt a purely mechanical standardization of procedures. Until such optima are firmly established, the procedures adopted with different techniques are likely to fall short of optimal conditions by different degrees. Since this is unlikely to affect all individuals equally, intercorrelations between measures will be correspondingly reduced.

2. Other Parameters of Conditioning Performance

The concept of "conditionability" refers to the ease of acquisition of conditional reflexes, i.e., the development of excitatory linkages. There are, however, other aspects of conditioning performance which are of relevance for personality functioning. Diamond et al. (1963) have argued cogently for their view that inhibitory processes enter into every behavioral act, and must be given equal consideration with excitatory processes if we are to understand any aspect of psychological functioning. The conditional reflex was developed by Pavlov as a methodological technique for studying the interaction of excitation and inhibition in higher nervous activity. Manifestations of inhibitory processes in conditioning were considered to occur in the phenomena of differentiation, the acquired capacity of a stimulus to suppress or delay a response to a long-acting stimulus, and extinction of conditional reflexes, both with and without reinforcement. These aspects of conditional reflex activity have received attention only occasionally in studies of conditioning and personality. An exception is to be found in the work of Solymon and Beach (1961), who, in their studies of memory defect in the aged, have investigated the conditioning parameters of rate of extinction, differen-

tiation, inhibition of delay, and rate of extinction of the orienting reflex, in addition to rate of acquisition of a conditional reflex.

It would seem that a concept of conditionability that deals only with excitatory processes ignores aspects of conditioning performance which, in the long run, could possibly account for more of the variance in personality functioning than excitatory measures such as rate of CR acquisition.

D. Problems and Possibilities of Specific Conditioning Techniques

1. Eyeblink

Conditioning of the eyeblink to a puff of air is a technique which has been widely used in personality study, and it is the most standardized method. It is standard practice to use of CS–US interval of 450–500 msec, and an air puff of approximately 1 psi. Also a pure tone is commonly used as the CS, mainly because light CS's have been shown to elicit UR's under some conditions. The transducer may be mechanical (microtorque potentiometer) or photoelectric. Both systems are highly developed and generally satisfactory. The different types of response which occur in eyeblink conditioning have been studied in detail. The most important criterion for distinguishing various types of response is latency, but form is also useful as a secondary criterion. The usually accepted categories of response and their latency ranges, with a 500-msec CS–US interval, are reflex blink to light (alpha response), 50–110 msec; beta response (reflex blink to light in dark-adapted eye), 120–240 msec; conditional response, 250–500 msec; voluntary response, 200–500 msec; unconditional response to puff, 50–100 msec after US.

Whilst UR's to the CS can readily be distinguished from CR's in terms of latency, the differentiation of voluntary responses from CR's poses a problem. The accepted criterion of a voluntary response is that of possessing characteristics similar to responses given by Ss who are instructed to blink to the CS in order to avoid the puff, or who report that they responded voluntarily in order to avoid the puff. Since the majority of these voluntary responses have latencies from 200 to 300 msec, Spence and Ross (1959) have proposed that only responses occurring in the interval 300–500 msec after CS onset should be termed CR's.

The form of a voluntary response is often characteristically different from that of a CR, but in many cases the criterion of form is difficult to apply, and there has been a tendency to rely on the criterion of latency.

Random blinking poses another problem for eyeblink conditioning, and some experimenters reject Ss whose natural blink rate is above a

predetermined level. Since a high blinking rate has often been considered a sign of anxiety, it is clear that the investigator of relations between personality and conditioning faces something of a dilemma in reaching a decision concerning whether or not to reject Ss. A similar problem arises in connection with Ss who respond voluntarily to the CS before training. It may happen that such voluntary responses are related to the personality parameter under study. For example, Willet (1963), in his study of the effect of high and low drive on eyeblink conditioning, rejected any S whose natural blink rate was faster than one blink every 500 msec. He reports that he rejected 7 of 62 Ss in the high drive group "through apparatus failure or because they responded to the CS signals during the control trials." Of the 60 low-drive Ss, on the other hand, "five were rejected through apparatus failure."

The personality psychologist's dilemma will be solved only by more fundamental work on the differentiation of associative and nonassociative factors in conditioning, and the relation of these factors to dimensions of personality. It will probably be necessary to recognize, however, that there are some Ss from whom adequate measures of eyeblink conditioning cannot be obtained, just as some Ss have an insufficiently defined alpha blocking response to permit EEG conditioning, and some Ss produce insufficient GSR's to make GSR conditioning possible.

2. Galvanic Skin Response

The galvanic skin response (GSR) rivals the eyeblink in popularity as a conditioning procedure with human Ss. The usual finding is that the GSR can readily be conditioned with an electric shock as the US, and it can also be conditioned with a loud auditory stimulus in place of the shock. An advantage of the GSR is that it varies continuously over a wide magnitude range, and hence lends itself to quantification. Nevertheless, there have been endless debates concerning the appropriate unit of measurement of the GSR. Investigators have laid down criteria for suitable measures, such as independence of basal resistance and normality of distribution. Units commonly employed have included percentage change in resistance, change in conductance, the square root of conductance change, and log conductance change. Various ratio measures have also been proposed. A large-scale study by Eysenck (1956) showed that the extent to which various measures met the usual criteria varied with the experimental conditions, and that none had any marked advantage over the others.

In the author's view, the problem of the most appropriate unit of measurement cannot be decided in the absence of an external criterion. Personality study offers the possibility of providing such a criterion. If,

for example, it were possible, on the basis of personality theory, to define groups with widely differing conditioning performance, these groups could be used as external criteria. That is to say, the GSR measure which most effectively and consistently separated samples from the groups in terms of conditioning would be the unit of choice. Until some such basis for a rational choice exists, it would seem reasonable to use the unit of measurement most convenient for one's purpose.

GSR conditioning is the least standardized of all the conditioning procedures. Whereas, in the case of eyeblink conditioning, the favored CS–US interval is almost always around 450–500 msec, this interval may vary from 450 msec to 20 sec in GSR conditioning. The characteristics of the US also vary considerably from experiment to experiment. Thus the electric shock US may differ in current type (for example, alternating or direct current, condenser discharge), electrode type, electrode placement, intensity, and duration. If the current type conforms to a standard pattern, for example 60 cycles ac, it can be specified adequately in terms of amperage if the electrode size and position are known. In many cases, however, the electric shock used to condition the GSR is unspecifiable.

The range of intertrial intervals commonly used in GSR conditioning is not great. The average is about 45–60 sec, but the interval may be as short as 15–25 sec (Prokasy, Hall, and Fawcett, 1962). The importance of the intertrial interval is emphasized by the failure of Prokasy et al. (1962) to find a difference between conditioning and pseudoconditioning groups with the very short interval of 15–25 sec, whereas separation of conditioning and pseudoconditioning groups was achieved with longer intertrial intervals in a later experiment (Wickens, Allen, and Hill, 1963).

What happens during the intertrial interval is another important source of difference between GSR experiments. The interval may be either "silent" or "filled." The term "silent" simply refers to the procedure, usual in other forms of conditioning, in which no other stimuli are presented between punctuate trials of CS, CS–US, or US. "Filled" (or "partially filled") intertrial intervals, on the other hand, are those in which a series of stimuli similar to the CS is presented between trials of CS, CS–US, or US. In the most common experiment of this type, the CS is a word, and other words are used to fill the intertrial interval (Becker, 1960; Becker and Matteson, 1961; Vogel, 1960, 1962). It can be seen that the "filled interval" experiment is simply one in which multiple differential stimuli are presented.

The question of filled versus unfilled intervals assumes some importance in the light of recent concern over the separation of associative from nonassociative factors in GSR conditioning. It has long been known that the GSR may be elicited by any novel stimulus, but electrophysio-

logical studies of the mechanisms of arousal and the orientation reflex have given this fact a new significance. The usual practice in GSR contioning is to present the stimulus to be conditioned a number of times prior to conditioning in order to adapt or habituate the GSR to this stimulus. After the adaptation trials, the conditioning phase is begun with several CS–US pairings, followed by the first test stimulus (CS presented alone). Any increase in the GSR to the test stimulus, beyond the level of response to that stimulus after adaptation, is attributed to conditioning. However, recent work by Sokolov (1960) and others has emphasized the extent to which the habituated orienting reflex, of which the GSR is a manifestation, is re-evoked by any relatively sudden change in the stimulus input. Evidence that the components of the orienting reflex can be habituated in different patterns, depending on the sensory input, suggests the operation of selective inhibitory mechanisms. To account for these and other findings, Sokolov (1960) has proposed the concept of "neuronal model." The neuronal model is a cortical representation, by a pattern of excitation, of recent past experience, against which incoming stimulation is compared. As long as the incoming stimulation corresponds to the model, it is checked by inhibition at the level of the brain stem collaterals, but in the case of a lack of concordance, afferent impulses are fed into the arousal system, which then discharges into cortical centers, increasing capacity for discrimination and causing the formation of a new model more consonant with the new stimulus input.

Grings (1960) has proposed a related concept which he terms the perceptual disparity reaction. Applied to the GSR conditioning situation, these concepts suggest, first, that the first reinforcement following an adaptation series will release the orienting reflex from inhibition, and this release will be reflected in subsequent response measures. In addition, moreover, any change in the pattern of stimulation, including the omission of a stimulus, will tend to disinhibit the orienting reflex. This means that occasional test trials of CS alone, in a series of CS–US pairings, will evoke an orientation reflex which will be confounded with any CR which occurs. If test trials are massed at the end of a conditioning sequence in the form of an extinction series, the initial trial will tend to evoke a strong orientation reflex, followed by the development of habituation on subsequent trials. Any conditional GSR's, and their subsequent experimental extinction, will be confounded with the orientation GSR's and their habituation.

It would be expected that the strength of the re-establishment of the orienting reflex by a change in the stimulus conditions will depend on the extent of the disparity between the original and the new stimulus

conditions, and the suddenness of the change. Consider, for example, the experiment of Prokasy *et al.* (1962) which compared the GSR's, in an extinction series, of groups given

(a) tone-shock pairings (regular conditioning),
(b) shock-tone pairings (backward conditioning),
(c) random shocks and random tones (sensitization),
(d) shocks only (pseudoconditioning), and
(e) tones only.

The ordering of the groups from highest to lowest GSR on the first extinction trial was shock only; backward conditioning; regular conditioning; random shocks and random tones; and tones only. This is precisely the ordering that judges give if they are asked to rank the groups in terms of decreasing perceptual disparity between training and extinction test trials. In the experiment of Prokasy *et al.*, the response levels of the shock-only and backward-conditioning groups decreased rapidly compared with those of the true conditioning group. The response level of the random group tended to remain relatively high, possibly as a result of trace conditioning with the very short intervals used.

Stewart, Stern, Winokur, and Fredman (1961) have argued recently that "so far work on GSR conditioning has dealt with the adaptation and recovery of unconditioned responses rather than conditioning of responses." The argument that *no* unequivocal evidence of GSR conditioning had previously been produced could be sustained only by ignoring the results of long-delay GSR conditioning experiments (Rodnick, 1937, Switzer, 1934), which measured anticipatory responses similar to those advocated by Stern *et al.*, and also the results of differential GSR conditioning experiments. The latter include experiments using a single differential stimulus (Kimmel and Pennypacker, 1963; Lovibond, 1963a), and those using multiple differential stimuli ("filled" intertrial intervals) (Becker, 1960; Becker and Matteson, 1961; Vogel, 1960, 1962).

Further unequivocal evidence of GSR conditioning has been obtained by experiments in which an extended extinction test series has demonstrated a significantly higher level of response in a forward conditioning group by comparison with a backward conditioning group (Prokasy *et al.*, 1962). The backward conditioning procedure provides an acceptable control, since the perceptual disparity between training and extinction is greater with this procedure than with forward conditioning.

From the foregoing analysis it would appear that incontrovertible evidence of GSR conditioning (other than differential conditioning) can most readily be obtained if the stimulus relations remain constant throughout the experiment, or if the changes in stimulus pattern are

below the level of discriminability of the "neuronal model" or orienting reflex comparator mechanism.

In practice, experiments of the first type are limited to those employing the delayed conditioning method in which anticipatory CR's are distinguished in terms of latency. It is necessary, of course, to employ a sufficiently long CS–US interval to avoid masking of anticipatory CR's by orienting GSR's. Conditioning with delays of 10 sec or longer presents no great difficulty, and contradicts the belief that GSR conditioning can be obtained only with a CS–US interval of 450–500 msec.

It seems likely that the function relating CS–US interval and strength of conditioning rises to a new peak beyond intervals of 4–5 sec, but curiously enough nobody has investigated this possibility. In considering the problem of CS–US interval, however, it should be borne in mind that measures of conditioning derived from short intervals will necessarily be confounded with orienting GSR's if silent intertrial intervals are used. This suggests that examination of the CS–US interval function can best be carried out with experiments which reduce the suddenness of changes in the stimulus pattern below the point where orienting GSR's will be evoked. This implies filling the intertrial interval with multiple differential stimuli to produce a "constancy in change." Current work in the Adelaide laboratories suggests that under these circumstances orienting GSR's will be almost entirely suppressed in many Ss. The fact that CR's appear to form readily enough against this background suggests that generalized arousal may not be necessary for the establishment of conditional linkages, and that the inhibition which suppresses the orienting reflex may have little or no effect on the elicitation of the GSR by specific arousal mechanisms. It must be stressed, however, that there is as yet insufficient evidence to permit firm conclusions concerning these possibilities.

From the point of view of personality study, the advantage of long-delay and filled-interval GSR conditioning procedures is that they give relatively pure measures of conditioning, and make possible a measure of rate of acquisition of the CR. Although the strength of orienting CR's is likely to be positively correlated with rate of conditioning (Lovibond, 1963a; Martin, 1963), confounding of the two measures is clearly undesirable if one wants to test theories relating personality and conditioning, or to investigate the generality of conditionability.

Positive reward conditioning of the GSR has been investigated by Lovibond (1963a,b) using still pictures of nude females as the US. Using a delayed differential conditioning procedure with silent intervals, it was found that the intrusion of orienting GSR's was much less pronounced than in the case of shock conditioning. Without distinguishing the la-

tencies of responses, a rising curve of response to the CS was obtained, together with a falling curve of response to the differential stimulus (DS). This technique promises to be a useful supplement to salivary conditioning in the investigation of positive reward conditioning. A "filled interval" moving film version is at present being investigated. In this form of the technique, short sequences of moving pictures of nudes are separated by intervals of neutral shots of outdoor scenery. Both visual and auditory long-delay conditional and differential stimuli are being used. Substitution of a shock US in this procedure should make possible a more satisfactory comparison of positive and negative conditioning of the GSR, since the usual confounding of orienting and conditional GSR's with aversive reinforcement should be markedly reduced.

A further response measure which might be used with this form of reinforcement is change in penis volume. Freund (1963) has reported on the use of a penile plethysmograph in the differentiation of homosexual from heterosexual interests. According to Freund (1962), change in penis volume is a more sensitive index of arousal by pictures of nudes than is the GSR. Since this response should be relatively independent of the generalized arousal or orienting response, its comparison with the GSR offers interesting possibilities in the investigation of generalized *versus* specific arousal in the formation of conditional reflexes.

3. Salivary Response

Since the principles of conditional reflex activity were derived by Pavlov from studies of salivary conditioning in animals, there would appear to be obvious advantages to be gained from the adoption of this method in the study of human conditioning. However, the salivary conditioning of human Ss poses some difficult methodological problems.

In early attempts to study salivary conditioning in man (Finesinger and Sutherland, 1939; Lentz, 1935; Richter and Wada, 1924), salivary secretion was measured by means of the suction cup method originally devised by Krasnogorski (1931). In the modified version of this method published by Finesinger and Finesinger (1937), capsules, held by suction over Stenson's ducts, collect the secretion of the parotid glands. The saliva collected is transmitted by rubber tubes to a bottle filled with saline solution. The displaced saline passes through a tube with a capillary tip into a burette. The passage of each drop of saline closes an electric circuit between two platinum wires in the wall of the burette, and sends a pulse to a kymograph pen. A measure of rate of salivary secretion is thus given by the number of pen movements per unit time. Lemon, or other fruit juice, which serves as the US, is conveyed to the patient's mouth via a third tube positioned by the suction cup.

In an experimental application of the technique, Finesinger *et al.* (1942) used a CS (metronome) presented for 3–10 sec in a delayed conditioning procedure, with a 1-sec CS–US overlap. The magnitude of the CR was measured by presenting the CS for 25 sec on test trials. The amount of saliva secreted during this period was corrected by subtraction of the average quarter-minute rate of secretion during the minute preceding the presentation of the test CS.

Razran (1935) published an account of a new technique for measuring salivary secretion in conditioning experiments with human Ss. Salivation was measured by means of a standard dental roll inserted under S's tongue. The roll was weighed before insertion and after removal, the difference in weight being taken as a measure of secretion during the intervening period. Razran noted considerable variation in the magnitude of the salivary CR, and believed that a good deal of this variation could be attributed to the influence of central "attitude" or "set" factors. When the attention of S was directed elsewhere, so that there was minimal awareness of the stimulus relations of the experiment, the stability of the CR improved (Razran, 1936).

The most recent studies of human salivary conditioning (Bindra *et al.*, 1955; Willet, 1960) have used the cotton roll method. Willet's carefully controlled study showed that increased salivation occurred in the presence of the CS when no reinforcement was used. The experiment of Bindra *et al.* was replicated in the Adelaide laboratories with the addition of a control group. The Ss in the control group were presented with the CS and US separated by random intervals rather than in conjunction. The curves of both experimental and control groups ($N = 20$ in each group) were virtually identical in form to those obtained by Bindra *et al.* On the first test trial, after 8 reinforcements, the amount of salivation had reached a maximum which was sustained through the next 32 CS–US pairings. The difference between the experimental and control groups was not significant, but the response level of the control group was consistently higher than that of the experimental group. In a comparable experiment, with a test trial after each reinforcement, most of the increase in salivation to the CS occurred after the first reinforcement, and no further increase occurred beyond three reinforcements. This pattern conforms closely to that observed in the responses of "sensitization" control groups used in GSR conditioning.

In a further unpublished study, John (1962) used a similar procedure with 80 experimental Ss and 80 control Ss. During the conditioning phase, salivation to the CS increased over 12 reinforced trials, but the rate of increase was significantly greater in the control group. A possible explanation of this result can be derived from Pavlov's (1927) observation

that if a CS of relatively long duration overlaps rather than precedes a relatively long US, the CS becomes inhibitory. John's Ss were instructed to start sucking a lollipop (US) on a given signal, and the CS and US were presented simultaneously for 2 min. Under these circumstances it is likely that the development of inhibitory potential by the CS offset the development of excitatory potential by this stimulus, by way of either sensitization or CS–US pairing.

The foregoing studies demonstrate quite unequivocally the occurrence of nonassociative contamination in the response measures of salivary conditioning experiments such as that of Bindra, but they do not elucidate the sources of this contamination. There are at least three possible explanations of the increased response to the CS during the conditioning phase observed in the control group Ss of the Adelaide and John experiments. The first two emphasize the possibility of measurement artefact, and the other suggests the occurrence of some form of pseudoconditioning.

First, when the intertrial interval is short, there may be a carryover of unconditional salivation into the period of CS test presentation. The intertrial interval in the experiments in question, as in the Bindra study, was only about 1 min. In this connection, Finesinger et al. (1942), using a suction cup method, have observed that sometimes salivation persists for "a considerable time" after presentation of the US. Second, presentation of the US (such as a lollipop) to a hungry person may sensitize the salivary reflex to the point where it will be elicited by any object (such as a cotton roll) placed in the mouth. Third, reinforcement may sensitize the salivary reflex to the previously indifferent CS, thus producing a form of nonassociative response increment similar to the sensitization that has been observed in aversive conditioning.

The possible contribution of each of these factors to response contamination in cotton roll salivary conditioning could readily be subjected to experimental analysis. If it is found important, the factor of carryover of salivation could be controlled by longer intertrial intervals. Sensitization would need to be controlled by such measures as the use of long-delay conditioning. As it is a reasonable inference from accounts of Pavlov's work that sensitization could scarcely have been a problem in work with dogs, it might seem a priori that it should not be a problem with human beings. There is, however, other evidence to suggest that the salivary reflex may be significantly more sensitive in human beings than in animals. For example, Finesinger et al. (1942) found that not a few Ss secrete saliva continuously irrespective of the external situation. Some degree of "interval" or intertrial secretion was so common that it was

necessary to correct for its effect on response measures. As Finesinger *et al.* point out, Pavlov makes no mention of interval secretion, and it is a reasonable inference that it was not a problem.

If, as seems likely, it is found that measurement artefact results from the use of cotton rolls themselves, there will be no choice but to return to the direct measurement of salivary secretion. This would appear highly desirable in any case, since direct measurement has a number of important advantages over the cotton roll technique. In the first place, continuous monitoring of salivary flow enables selection of the intertrial interval necessary to avoid carryover of unconditional salivation. Second, it gives a precise measure of interval secretion, and thus makes possible correction for background level of salivation. Third, the method of direct presentation of the US, which is a natural corollary of direct recording, permits far more precise control over the timing and amount of reinforcement than can be achieved with the cotton roll method. Fourth, the whole procedure can be carried out without any movement whatever on the part of S. Once S has adapted to the tubes and suction cups in his mouth, the course of conditioning is uninterrupted by any extraneous stimuli. In other words, we have a situation similar to that which Pavlov found essential for the salivary conditioning of dogs. By contrast, the cotton roll method , as it is commonly used, requires S to insert some solid reinforcing agent (lollipop or tablet) into his mouth, and to begin sucking on a given signal. In addition, S not infrequently has the US in view and is required to engage in some task. It seems not unlikely that these varied patterns of stimulus input will either facilitate or inhibit salivary flow, and account for at least some of the variability of response which is commonly noted. The results obtained with the suction cup method make it clear, nevertheless, that there is a good deal of inherent variability in the human salivary response. A direct experimental comparison of the suction cup and cotton roll methods would be most instructive.

It is of interest to consider possible modifications to the Finesinger and Finesinger suction cup method in the light of recent developments in instrumentation. Possibilities which suggest themselves include (a) substitution of an inkwriting rate recorder for the kymograph; (b) presentation of the liquid US by a small, fixed-delivery pump, automatically timed; (c) substitution of a plastic cannula for the suction cup. It is not immediately obvious what improvement could be made on the original method of transducing salivary flow into electrical impulses, but it is quite likely that a superior method exists or could be developed. It is clear that, modern developments notwithstanding, direct salivary record-

ing methods are technically demanding. Nevertheless, the available evidence suggests that serious application of salivary conditioning procedures in the study of personality can scarcely be undertaken without them.

4. Alpha-Blocking Response

The early work of Durup and Fessard (1935) and Loomis, Harvey, and Hobart (1936) suggested the possibility of conditioning the blocking or desynchronization of the alpha rhythm, for which a strong light is an adequate stimulus.

In most experiments a tone has been used as the conditional stimulus. An extensive exploratory study was undertaken by Jasper and Shagass (1941) to (a) investigate "the depression (blocking) of the alpha rhythm in man as a conditional response, and (b) appraise the value of the electroencephalograph as an instrument for the study of conditioning problems, especially in man." Jasper and Shagass claimed to have demonstrated practically all of the Pavlovian conditioning phenomena, including temporal, trace, delayed, differential, and backward conditioning. In their view, the advtantages of EEG conditioning lie in the facility and rapidity with which conditioning of alpha blocking may be achieved, and its relative independence of voluntary influence. The main disadvantages are the instability of the CR and the fact that S must exhibit "an almost continuous alpha rhythm which is free from 'spontaneous' variations, and which responds clearly to visual stimuli" before unequivocal conditional responses can be obtained.

Most investigators who have studied EEG conditioning in man have found the CR to be "poorly sustained" (Wells and Wolff, 1960a) or "unstable" (Knott and Henry, 1941). Knott and Henry (1941), in a very careful study, drew attention to the possibility of sensitization of the response to the tone in EEG conditioning. They noted that alpha blocking can be produced by a wide variety of stimuli. In fact, the tones ordinarily used as conditional stimuli are by no means "indifferent" stimuli; they are only relatively ineffective in comparison with strong light stimuli. In order to differentiate between sensitization of the previously adapted response to the conditional stimulus, and conditioning proper, Knott and Henry used the method of delayed conditioning with a 4-sec CS–US interval. This interval permitted full recovery of the alpha rhythm after its response to the CS and before the US occurred. The results gave rather clear-cut evidence of both sensitization of the initial response to the CS and the development of a conditional anticipatory response in the 1-sec interval prior to US onset.

The curve of the sensitized response to CS onset showed an abrupt rise to approach its maximum during the first five trials, although the

peak did not occur until trials 21–30. From this point there was a slow decline in the response. The anticipatory response showed a negatively accelerated increase to reach its peak in trials 21–30, followed by a fairly rapid decline.

Despite this early demonstration that nonassociative factors such as sensitization must be controlled in EEG conditioning, not all subsequent investigators have heeded the warning. Thus the Wells and Wolff (1960a) study of EEG conditioning in brain-damaged patients employed a CS–US interval of only .8 to 1 sec, and did not include a control group. A temporary cerebral connection was considered to have occurred "if the alpha rhythm was obliterated or strikingly depressed after the presentation of the tone and before the appearance of the light." The "conditioning" curve obtained showed a decline from an initial peak value similar to the usual adaptation curve. Stern, Das, Anderson, Biddy, and Surphlis (1961) questioned whether Wells and Wolff had demonstrated the development of a conditional cerebral response, and undertook a replication of the Wells and Wolff experiment. Their results showed a decline in the desynchronization response to tone over the 20 adaptation trials, followed by a rise to above the preadaption level in the first five trials of tone-light pairing, and a subsequent rather regular decline over the next 40 trials. Since the curve of the response to the tone almost exactly paralleled that of the response, to the light, Stern *et al.* interpret the rise in tone response, during the early stage of tone-light pairing, to a return of the adapted "orienting response" analogous to that observed in GSR conditioning.

The implication of these results is that Wells and Wolff did not provide unequivocal evidence of EEG conditioning, and hence, did not provide evidence concerning the differential susceptibility to this type of conditioning of brain-damaged patients and normal controls. Since Wells and Wolff (1960b) used an identical procedure in their study of anxiety and EEG conditioning, it follows that this study too failed to produce evidence of differential rates of conditioning in anxious patients and normals.

It should be obvious that Wells and Wolff have simply failed to disprove the null hypothesis of no difference between their clinical groups and normals. It is probable, however, that there is a positive correlation between disinhibition of adaptation and rate of conditioning [as Lovibond (1963a) has demonstrated for positive reward conditioning of the GSR] and that results similar to those of Wells and Wolff would be obtained with unequivocal measures of EEG conditioning.

It is clear from the foregoing that conditioning of alpha desynchronization is susceptible to masking by nonassociative factors in much

the same way as the conditioning of other components of the orienting reflex and also the defensive reflex. When it is considered that only about 60% of normals produce EEG records with alpha rhythm present 50% of the time (Davis, 1938; Henry and Knott, 1941), it is clear that conditioning of alpha blocking leaves much to be desired as a method of personality study.

Nebylitsyn (1963) describes a method which he claims improves the stability of the conditional reflex connection in EEG conditioning. He notes that the reaction of depression or complete disappearance of the alpha rhythm always occurs upon the presentation of a light stimulus, and, under conditions of "novelty," when a stimulus of any modality appears. "In the former instance this reaction possesses virtually all the properties of an unconditional reflex, while in the second case it possesses those of an orientation reflex" (Nebylitsyn, 1963, p. 23). In subsequent discussion, however, Nebylitsyn points out that "rapid and significant diminution" of the "unconditional" depression of the alpha rhythm usually occurs with repeated presentation of the light stimulus. Consequently, he uses what he terms "activating reinforcement" in the elaboration of conditional blocking of the alpha rhythm. Activating reinforcement refers to a reinforcement that "supports the orientation-investigatory reaction of the individual and thereby makes for stable retention of the conditional reflex." A series of slide pictures displayed on a screen is employed for the purpose of activating reinforcement, and Nebylitsyn (1962) claims that more than 90% of Ss with a reasonably clear-cut alpha rhythm develop stable CR's using this method.

IV. Changing Personality Functioning by Conditioning Methods

A. Experimental Neuroses

One of the earliest applications of conditioning procedures to problems of personality was the study of experimental neuroses.

This line of research began with some observations made in Pavlov's laboratory while training a dog in a difficult differentiation (Pavlov, 1927). After a CR of salivation to a circle was formed, a series of ellipses with different semiaxis ratios was presented as differential stimuli. Following the usual practice, simple differentiation was first established by reinforcing the circle and omitting reinforcement to an ellipse with a ratio of 2:1 between the semiaxes.

In following trials the shape of the ellipse was changed by steps until it approximated that of the circle. When an ellipse with a ratio between the semiaxes of 9:8 was presented the dog's behavior suddenly changed markedly. Its capacity for differentiation deteriorated, and its general

behavior changed from quiet cooperation to violent resistance. Pavlov used the term "experimental neurosis" to characterize the sudden disturbance in behavior which developed under the given experimental conditions.

In another experiment, a salivary CR to a strong electric shock was established, but when an attempt was made to generalize the CR by applying the shock to other parts of the skin, a breakdown of behavior occurred. The salivary CR disappeared and was replaced by a violent defensive reaction even in response to weak stimuli. Pavlov interpreted these observations as demonstrating that the clash of excitation with inhibition led to a profound disturbance of the usual balance between these two processes "and finally gave way to an undisputed predominance of one of them, producing a pathological state."

A comprehensive and detailed review of later work on experimental neuroses (Liddell, 1947, 1956; Gantt, 1942, 1944; Masserman, 1943; Wolpe, 1952, 1958) has been published recently by Broadhurst (1960). Consequently the present treatment will be confined to a brief discussion of some of the issues of general importance which have arisen in relation to this work.

Broadhurst very laudably sets out to establish criteria for determining whether or not experimentally induced behavioral disturbances in animals may properly be termed neuroses. In this endeavor he is guided by Hebb (1947), who lists the following six criteria of human neuroses: The behavior must be (a) undesirable, or evaluatively abnormal, (b) emotional, (c) generalized, i.e., not solely a response to a specific excitant, (d) persistent following cessation of the specific excitant, (e) statistically abnormal, i.e., occurring in a minority of the population, (f) without origin in a gross neural lesion. Broadhurst, with Hebb, rejects the following three common criteria: (i) neurosis has no known physiological base, (ii) neurosis produces a marked change of behavior from early base line, (iii) neurosis follows from some "traumatic" experience like conflict or frustration.

Broadhurst very properly concludes that a great deal of the work in the field has lacked rigor and sophistication in experimental design, and has been inadequately reported. Furthermore, many of the behavior disturbances produced fall far short of warranting the description experimental neurosis. However, Broadhurst meets certain difficulties in the application of Hebb's criteria. Because of lack of data concerning the behavior of experimentally neurotic dogs external to the experimental situation, he is forced to conclude that there is insufficient proof that the breakdowns of behavior produced in Pavlov's laboratories can properly be described as neuroses. This difficulty arises, of course, from the fact that Pavlov himself emphasized criteria not included in Hebb's list,

particularly the criterion of disorders in conditional reflex activity other than that activity through which the disturbance was produced. An example was the failure of the dog in the original experiment to perform even the simplest differentiation which had previously been performed with ease. Other examples were the disappearance of all recently trained reflexes, either inhibitory or excitatory, and performing a negative response to a positive stimulus, and vice versa. For Pavlov these were the essential criteria, and it is of interest that nothing like them occurs in Hebb's list. The closest approach to this sort of criterion would be one of those rejected by Hebb, namely, "neurosis produces a marked change in behavior from an earlier baseline." Certainly, as it is formulated by Hebb, this criterion would fail to distinguish the behavior under consideration from a number of nonneurotic behavioral changes, but it would seem to do so if reformulated along the following lines: "inability to perform tasks which were performed with ease prior to breakdown, or which are performed with ease by individuals of similar behavioral capacity." This would seem to be a not unimportant criterion of human neurosis. If a patient with an hysterical paralysis is unable to lace his shoes, a practiced accountant is unable to add a few simple figures, or a dweller in a large city is unable to enter and operate an automatic elevator, we have examples of this sort of behavioral deficiency.

Broadhurst rejects the various theoretical interpretations of experimental neuroses which have been made previously, and seeks to account for the phenomena in terms of conditional emotional responses. Both Hebb (1947) and Wolpe (1952) earlier offered a similar interpretation. Hebb was forced to make an exception in the case of breakdown occasioned by training in very difficult differentiation as originally reported by Pavlov. Wolpe, however, in support of his thesis that all experimental neuroses, and all human neuroses, can be regarded as maladaptive learned behavior patterns, attempts to bring the difficult differentiation case within his general framework. Wolpe classifies situations which precipitate experimental neuroses into those which use strong shock, those which use mild shock, and those which present ambiguous stimuli. Experiments of the first type produce conditional fear or anxiety, i.e., stimuli associated with the strongly aversive shock come to elicit intense anxiety by virtue of this association. (It is of interest to note that the Pavlovians draw a sharp distinction between experimental neuroses in which there is evidence of generally disordered nervous activity as previously discussed, and manifestations of "defense reflex" activity unaccompanied by signs of general disorder.)

In order to explain how the repeated application of mild shock can eventually give rise to a severe anxiety reaction, Wolpe postulates that

"when emotional conditioning first begins to be established, the conditioned stimulus evokes a small amount of anxiety. Let us call this x, and let it be half the amount evoked by the shock $(2x)$. At this time, when the shock follows shortly on the conditioned stimulus, the anxiety from the two sources will add together, giving a total of $3x$. As the result of conditioning, the next presentation of the conditioned stimulus will evoke, say, $1\frac{1}{2}x$, and the shock that follows will bring the total to $3\frac{1}{2}x$. It may then be found that the next presentation of the conditioned stimulus evokes $2\frac{1}{4}x$, that is, more than the shock itself. By repetition, the amount of anxiety evocable by the conditioned stimulus will gradually rise to a magnitude many times $2x$. At some stage the anxiety will be great enough to produce obvious signs, and as experimenting continues these will be intensified."

Wolpe's interpretation of the third precipitating situation, that of "ambiguous stimulation," is as follows. However slight and transient the anxiety produced by this conflict of tendencies may be at first, the drive reduction associated with its cessation results in the anxiety responses becoming reinforced to whatever stimuli are contiguous, and among these is the ambivalent stimulation itself. In the same way as suggested above for the case of mild noxious stimuli, the strength of the anxiety response is conceived to be gradually stepped up at each presentation of the ambivalent stimulus situation while the drive-reduction potential correspondly grows. Eventually very powerful anxiety responses are evoked, and strong avoidance behavior entirely replaces the approach responses to the experimental situation which were previously established by repeated feeding (Wolpe, 1952, p. 266).

Broadhurst does not regard Wolpe's interpretation as wholly satisfactory, and he suggests an alternative way of bringing the behavior disturbance resulting from ambiguous stimulation into the general category of conditional emotional reactions. Broadhurst's hypothesis involves the application of a reformulation of the Yerkes–Dodson law. According to this law, there is a curvilinear relationship between motivation and performance. Performance improves with increased motivation up to an optimum, beyond which further increase in motivation results in poorer performance. Second, the more difficult the task the lower the optimum level of motivation. In Hullian terms, "increasing drive level may increase the functional difficulty of a discrimination, so that objective increases in difficulty are in fact rendered more severe than the circumstances otherwise warrant. Once the optimum motivation is reached, then the decrease in task performance may be viewed as an energization of incorrect responses which are lower in the habit family hierarchy than the correct ones" (Broadhurst, 1960, p. 754).

Broadhurst's application of these principles to the difficult differ-entiation situation is as follows. Initially S's fear reactions to the restraint and to the experimental situation generally are gradually inhibited and replaced by food-seeking responses. In training the discrimination, the difficulty of the task progressively increases. Optimum performance would thus require a progressive lowering of motivation, but as S is likely in the past to have encountered mainly simple situations requiring high drive for optimum performance, he is likely to have learned to respond to failure by increased motivation. Failure to receive reinforcement for wrong response to the negative stimulus thus results in increased drive, resulting in further failure and so on. Broadhurst suggests that at a certain stage "the increased drive level energizes incorrect responses in competi-tion with the correct ones—responses, especially fear ones, which were suppressed early on in the training. So it is that struggling, howling, and agitated behavior in general, which were previously characteristic of the subject's initial reactions to the sitaution are reinstated anew" (Broad-hurst, 1960, p. 755).

Finally, in relation to the finding that after the behavior disturbance occurs S is unable to make even simple differentiations, Broadhurst says, "It seems probable that this characteristic failure of a simple discrimina-tion can be accounted for if it is recalled that even relatively simple tasks have an optimum motivation, which may well be surpassed if anxiety is added to the motivational complex."

The attempts of Wolpe and Broadhurst to account for the phenom-ena of difficult differentiation neuroses in terms of learned emotional reactions, although ingenious, are scarcely convincing. Wolpe's assump-tion of an additive relationship between conditional and unconditional fear is somewhat implausible, especially in relation to breakdown in the difficult differentiation experiment. Equally implausible is Broadhurst's hypothesis that increased drive level energizes initial fear reactions to produce the "struggling, howling and agitated behavior" observed when breakdown occurs. In the first place, it would appear from the descriptions available that animals in whom breakdown occurred did not necessarily show initial fear or resistance to the restraining harness. Broadhurst's hypothesis that increasing drive results from continued failure, and in-creases the probability of further failure, is, in all probability, correct, but it is unlikely that the increased drive is dependent on prior learning. It is far more likely that the increase is an unconditional effect of failure of stimulus input to match the "neuronal model" or "expected" input (Sokolov, 1960). Broadhurst's increasing-drive hypothesis thus gives a plausible account of why discrimination fails, but not of the resultant breakdown of behavior. One aspect of behavior disturbance in discrimina-

tion experiments which creates difficulties for both theories is the relative suddenness with which the behavior disturbance typically occurs. The term "breakdown" appears to give an apt description of what happens. The theories of both Wolpe and Broadhurst, however, would lead one to expect a gradual development of behavioral disturbance. As it is also difficult to see how either theory could account for the many phenomena of disturbance of "phasic activity" which Pavlov (1927) describes, there would seem to be reason to question whether the phenomena of experimental neuroses in animals can adequately be described purely in terms of maladaptive responses. This question has important implications for the therapy of human neuroses, which will be taken up briefly in section IV, B.

B. Therapy of Behavior Disturbances by Direct Conditioning Methods

During the past decade there has been an increasing interest in the possibility of treating behavior disorders by direct conditioning methods. The procedures used are direct in the sense that they are aimed at the elimination of undesirable behavior patterns, or the development of desirable patterns, by the presentation of specific and precise stimulus relationships. By contrast, the conventional therapeutic methods based on Freudian theory, or one of its derivatives, may be termed indirect because attention is directed primarily toward the "complex" which is assumed to give rise to the symptoms.

The method of direct conditioning in the treatment of behavior disorders has been termed "behavior therapy" by Eysenck (1960), but not all psychologists who use direct conditioning techniques find the implications of this term acceptable. Eysenck (1960) has brought together virtually all of the papers dealing with behavior therapy which were published prior to 1960. The disorders which have been treated by this method with apparent success include phobias, obsessions, stuttering, fetishes, writer's cramp, alcoholism, hysterical paralyses, frequency of micturition, and enuresis.

Criticisms of direct conditioning methods of treatment have centered mainly around the charge that the treatment deals only with symptoms, and since the underlying conflict is left unresolved, symptom substitution or exacerbated anxiety is to be expected. No clear case of symptom substitution appears to have been recorded, however, and the general personality changes accompanying the relief of specific symptoms appear to be beneficial in the great majority of cases (Eysenck, 1960; Lovibond, 1964b).

Studies of direct conditioning treatment have usually dealt with

single cases, or at best very small numbers of Ss, and hence are open to the criticism that, as in the case of verbal psychotherapy (Eysenck, 1960), improvement over spontaneous recovery rates has not been demonstrated. In a few cases, predictions of the outcome of changes in therapeutic technique (e.g., Yates, 1958) have provided strong evidence of internal consistency, but only controlled group studies can establish the efficacy of direct conditioning treatment beyond any doubt. Eysenck (1960) has discussed the problems involved in providing unequivocal evidence of the effectiveness of any form of treatment of neurotic disorders. In the case of enuresis, the problems are far less formidable, and there is now sufficient evidence to warrant making direct conditioning therapy the treatment of choice for this disorder (Jones, 1960; Lovibond, 1964b). Initial arrest of wetting is achieved in 90% of cases, but as in direct conditioning treatment generally, the problem of preventing relapse or extinction of conditioning is a pressing one. Recent experimental work has been directed to the application of laboratory techniques, including partial reinforcement schedules, to this problem (Lovibond, 1963c). By and large the demonstrational phase of direct conditioning treatment has passed, and it can be expected that the future will see more research into the precise mechanisms of particular behavioral disturbances, and a consequent increase in the precision of application of conditioning principles.

The limits of applicability of direct conditioning methods is a moot question. We have seen that theorists such as Wolpe (1952, 1958) regard all neurotic disorders as learned maladaptive behavior patterns, and hence susceptible to direct conditioning or behavior therapy. According to this approach, which has been discussed in detail by Eysenck (1960) and Yates (1958), a neurosis is symptoms and nothing else. It follows that treatment can only be symptomatic, i.e., aimed directly at the maladaptive behavior patterns. This conception of neurotic disorders is usually contrasted with the Freudian or "dynamic" view that the neurosis is an unconscious conflict of which the symptoms are merely surface indicators.

There is, however, a third view which has been associated in particular with the work of Pavlov (1927) and Hebb (1947). According to this view, it is necessary to make a sharp distinction between behavior arising from a central state or event and that central state itself, in the present case between neurotic behavior and an hypothesized disturbance of nervous activity which constitutes the neurosis. This, of course, is a distinction which those who take the "maladaptive habit" view of neurosis do not recognize.

If the distinction is accepted, a satisfactory explanation of neuroses must be sought in the neurophysiological mechanisms of behavior rather than the behavior itself. The implication of this point of view is that

neurosis may be manifest primarily in a set of maladaptive behavior patterns, or it may be manifest primarily in an inability to cope with previously simple tasks and problems, and difficulty in acquiring new, adaptive learning. To the extent that the latter features predominate, there will be an emphasis in therapy on procedures designed to affect the central state as such, including rest and sleep therapy, and the use of drugs, but conditioning procedures may be used in the development of new adaptive behavior patterns. To the extent that maladaptive behavior patterns predominate, the therapeutic emphasis may well be placed on such methods as reciprocal inhibition, desensitization, and aversive therapy. Those maladaptive behavior patterns which have been acquired under conditions of trauma or extremes of excitation are likely to have the character of S-R linkages, which, it is reasonable to hypothesize, will be outside the range of possible central integration and modification. It follows that the behavior patterns in question will be beyond the scope of verbal psychotherapy, and direct conditioning procedures will be the only methods with a reasonable chance of success. It will be recognized, however, that neurosis will often involve disturbances of central processes as manifest in attitudes and beliefs, for the modification of which verbal psychotherapeutic procedures will be appropriate. The form of such psychotherapy will not necessarily bear a very close resemblance to conventional Freudian procedures.

References

Al-Issa, I. The effect of attitudinal factors on conditioning. In H. J. Eysenck (Ed.), *Experiments in motivation.* New York: Pergamon Press, 1963.

Anand, B. K., and Brobeck, J. R. Hypothalamic control of food intake in rats and cats. *Yale J. Biol. & Med.,* 1951, **24,** 123-140.

Anokhin, P. K. Electroencephalographic analysis of corticosubcortical relations in positive and negative conditioned reactions. *N.Y. Acad. Sci.,* 1961, **92,** Art. 3, 899-938.

Barendregt, J. P. *Research in psychodiagnostics.* The Hague, Netherlands: Mouton, 1961.

Baron, M. R., and Connor, J. P. Eyelid conditioned responses with various levels of anxiety. *J. exp. Psychol.,* 1960, **60,** 310-313.

Beam, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.,* 1955, **51,** 543-551.

Becker, W. C. Cortical inhibition and extraversion-introversion. *J. abnorm. soc. Psychol.,* 1960, **61,** 52-66.

Becker, W. C., and Matteson, H. H. GSR conditioning, anxiety, and extraversion. *J. abnorm. soc. Psychol.,* 1961, **62,** 427-430.

Bender, L., and Schilder, P. Unconditioned and conditioned reactions to pain in schizophrenia. *Amer. J. Psychiat.,* 1930, **10,** 365-384.

Bindra, D., Paterson, A. L., and Strzelecki, J. On the relation between anxiety and conditioning. *Canad. J. Psychol.,* 1955, **9,** 1-9.

Bitterman, M. E., and Holtzman, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615-623.

Broadhurst, P. Abnormal animal behavior. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. London: Pitman, 1960, pp. 726-763.

Cook, S. W., and Harris, R. E. The verbal conditioning of the Galvanic Skin Reflex. *J. exp. Psychol.*, 1937, **21**, 202-210.

Das, J. P. An experimental study of the relation between hypnosis, conditioning and reactive inhibition. Thesis submitted for Ph.D. Univer. of London, 1957.

Davis, H. The electroencephalogram. *Tabulae Biologicae*, 1938, **16**, 116-131.

Davis, R. C., Buchwald, A. M., and Frankmann, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychol. Monogr.*, 1955, No. 405, **69** (20).

Diamond, S., Balvin, R. S., and Diamond, F. R. *Inhibition and choice: A nuero-behavioral approach to problems of plasticity in behavior*. New York: Harper, 1963.

Durup, G., and Fessard, A. L'éctrencéphelogramme de l'homme. *L'annee Psychol.*, 1935, **36**, 1-32.

Eysenck, H. J. *The dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.

Eysenck, H. J. *Handbook of abnormal psychology*. London: Pitman, 1960.

Eysenck, H. J. Conditioning and personality. *Brit. J. Psychol.*, 1962, **53**, 299-305.

Eysenck, H. J., and Claridge, G. The position of hysterics and dysthymics in a two-dimensional framework of personality description. *J. abnorm. soc. Psychol.*, 1962, **64**, 46-55.

Eysenck, S. B. G. An experimental study of psychogalvanic reflex responses of normal neurotic, and psychotic subjects. *J. Psychosom. Res.*, 1956, **1**, 258-272.

Field, I. G., and Brengelmann, J. C. Eyelid conditioning and three personality parameters. *J. abnor. soc. Psychol.*, 1961, **63**, 517-523.

Finesinger, J. E., and Finesinger, G. L. A modification of the Krasnogorski method for stimulating and measuring the secretions from the parotid glands in human beings. *J. lab. clin. Med.* 1937, **23**, 267-273.

Finesinger, J. E., and Sutherland, G. F. The salivary conditional reflex in man. *Trans. am. neurol. assn.*, 1939, p. 50.

Finesinger, J. E., Sutherland, G. F., and McGuire, F. F. The positive conditional salivary reflex in psychoneurotic patients. *Amer. J. Psychiat.*, 1942, **99**, 61-74.

Fitzwater, M. E., and Thrush, R. S. Acquisition of a conditioned response as a function of forward temporal contiguity. *J. exp. Psychol.*, 1956, **51**, 59-61.

Franks, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143-150.

Franks, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, **48**, 119-126.

Franks, C. M. Conditioning and abnormal behavior. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. London: Pitman, 1960.

Freund, K. Personal communication, 1962.

Freund, K. A laboratory method for diagnosing predominance of homo- or hetero-sexual interest in the male. *Behav. Res. Ther.*, 1963, **1**, 85-93.

Gantt, W. H. The origin and development of nervous disturbances experimentally produced. *Amer. J. Psychiat.*, 1942, **98**, 475-481.

Gantt, W. H. *Experimental basis for neurotic behavior: Origin and development of artificially produced disturbances of behavior in dogs*. New York: Hoeber-Harper, 1944.

Gilberstadt, H., and Davenport, G. Some relationships between GSR conditioning and judgments of anxiety. *J. abnorm. soc. Psychol.*, 1960, **60**, 441-443.

Grings, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **67**, 243-252.

Guk, E. D. The conditioned reflex activity in schizophrenics. *Sovetsk. Nevropatol.*, 1934, pp. 76-84.

Hebb, D. O. Spontaneous neurosis in chimpanzees. Theoretical relations with clinical and experimental phenomena. *Psychosom. Med.*, 1947, **9**, 3-16.

Henry, W. E., and Knott, J. R. A note on the relationship between personality and the alpha rhythm of the electroencephalogram. *J. exp. Psychol.*, 1941, **28**, 362-366.

Hilgard, E. R., and Humphreys, L. G. The effect of supporting and antagonistic voluntary instructions on conditioned discrimination. *J. exp. Psychol.*, 1938, **22**, 291-304.

Howe, E. S. GSR conditioning in anxiety states, normals, and chronic functional schizophrenic subjects. *J. abnorm. soc. Psychol.*, 1958, **56**, 183-189.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Hull, C. L. *A behavior system.* New Haven, Connecticut: Yale Univ. Press, 1952.

Ivanov-Smolensky, A. G. Uber die bedingten Reflexe in der depressiven Phase das manisch-depressiven Irreseins. *Mschr. Psychiat. Neurol.*, 1925, **58**, 376-388.

Jasper, H., and Shagass, C. Conditioning the occipital alpha rhythm in man. *J. exp. Psychol.*, 1941, **28**, 373-388.

Jenkins, J. J., and Lykken, D. Individual differences. *Ann. Rev. Psychol.*, 1957, **8**, 79-112.

Jensen, A. Personality. *Ann. Rev. Psychol.*, 1958, **9**, 295-322.

John, I. D. Personal communication. Univer. of Melbourne, Australia, 1962.

Jones, H. G. The behavioral treatment of enuresis nocturna. In H. J. Eysenck (Ed.), *Behavior therapy and the neuroses.* New York: Pergamon Press, 1960.

Kaada, B. R., and Johannsen, N. B. Generalized electrocortical activation by cortical stimulation in the cat. *EEG clin. Neurophysiol.*, 1960, **12**, 567-573.

Kimble, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1-15.

Kimble, G. A. *Hilgard and Marquis' Conditioning and learning.* (2nd ed.) New York: Appleton-Century-Crofts, 1961.

Kimble, G. A., Mann, L., and Dufort, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49**, 407-417.

Kimmel, H. D., and Pennypacker, H. S. Differential GSR conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1963, **65**, 559-563.

Knott, J. R., and Henry, W. E. The conditioning of the blocking of the alpha rhythm of the human electroencephalogram. *J. exp. Psychol.*, 1941, **28**, 134-144.

Krasnogorski, N. I. Bedingte und unbedingte Reflexe im Kindersalter und ihre Bedeutung für die Klinik. *Ergebn. d. inn. Med. u. Kinderh.*, 1931, **36**, 613.

Lacey, J. I., Smith, R. L., and Green, A. Use of conditioned autonomic responses in the study of anxiety. *Psychosom. Med.*, 1955, **17**, 208-217.

Landkof, B. L. Unconditioned and conditioned vascular reflexes in schizophrenics. *Trud. Tsentral. Psikhonevrol. Inst.*, 1938, **10**, 37-63.

Lentz, A. K. Les reflexes conditionnels salivaires chez l'homme sain et aliéné et leur rapprochement avec les donnes de la conscience. *L'encephale*, 1935, **11**, 394.

Liddell, H. S. The experimental neurosis. *Ann. Rev. Physiol.* 1947, **9**, 569-580.

Liddell, H. S. *Emotional hazards in animals and man.* Springfield, Illinois: Charles C Thomas, 1956.

Loomis, A. L., Harvey, E. N., and Hobart, G. A. Brain potentials during hypnosis. *Science*, 1936, **83**, 239-241.

Lovibond, S. H. The Object Sorting Test and conceptual thinking in schizophrenia. *Austral. J. Psychol.*, 1954, **6**, 52-70.

Lovibond, S. H. Conceptual thinking, personality, and conditioning. *Brit. J. soc. clin. Psychol.*, 1963a, **2**, 100-111.

Lovibond, S. H. Positive and negative conditioning of the GSR. *Acta Psychologica*, 1963b, **21**, 100-107.

Lovibond, S. H. Intermittent reinforcement in behavior therapy. *Behav. Res. Ther.*, 1963c, **1**, 127-132.

Lovibond, S. H. Manual for the Object Sorting Test. Unpublished manuscript, Univer. of Adelaide, Australia, 1964. (a)

Lovibond, S. H. *Conditioning and enuresis*. New York: Pergamon Press, 1964. (b)

McAllister, W. R. Eyelid conditioning as a function of the CS-US interval. *J. exp. Psychol.*, 1953, **45**, 417-422.

Malmo, R. B. Measurement of drive: an unsolved problem in psychology. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation*. Lincoln: Univer. of Nebraska Press, 1958.

Martin, I. GSR conditioning and pseudoconditioning. *Brit. J. Psychol.*, 1962, **53**, 365-371.

Martin, I. Delayed GSR conditioning and the effect of electrode placement on measures of skin resistance. *J. Psychosom. Res.*, 1963, **7**, 15-22.

Masserman, J. H. *Behavior and neurosis: an experimental psychoanalytic approach to psychobiologic principles*. Chicago, Illinois: Univer. of Chicago Press, 1943.

Mays, L. L. Studies of catatonia. V. Perseverational tendencies in catatonic patients. *Psychiat. Quart.*, 1934, **8**, 728-735.

Meehl, P. E. Schizotaxia, schizotypy, and schizophrenia. *Amer. Psychologist*, 1962, **17**, 827-838.

Moeller, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162-166.

Moruzzi, G., and Magoun, H. W. Brainstem reticular formation and activation of the EEG. *EEG clin. Neurophysiol.*, 1949, **1**, 455-473.

Nebylitsyn, V. D. Die Anwendung der Faktorenanalyse bei der Erforschung der Struktur der höheren Nerventätigkeit. *Probleme und Ergebnisse der Psychologie*, 1961, 3-4, 119-126.

Nebylitsyn, V. D. Personal communication, 1962.

Nebylitsyn, V. D. Certain electroencephalographic indicators of equilibrium in nervous processes. *Soviet Psychol. & Psychiat.*, 1963, **1**, 22-27.

Nebylitsyn, V. D., Rozhdestvenskaya, V. I., and Teplov, B. M. Concerning the inter-relation between absolute sensitivity and strength of the nervous system. *Quart. J. exp. Psychol.*, 1960, **12**, 17-25.

O'Connor, N. 1959. Ref. in Willet, R. A. Measures of learning and conditioning. In H. J. Eysenck (Ed.), *Experiments in personality*. London: Routledge & Kegan Paul, 1960, Vol. II.

Olds, J. A preliminary mapping of electrical reinforcing effects in the rat brain. *J. comp. physiol. Psychol.*, 1956, **49**, 281-285.

Olds, J. Effects of hunger and male sex hormone on self-stimulation of the brain. *J. comp. physiol. Psychol.*, 1958, **51**, 320-324.

Olson, W. C. *Problem tendencies in children*. Minneapolis: Univer. of Minnesota Press, 1930.

Pavlov, I. P. *Conditioned reflexes*. (Transl. by C. V. Anrep.) London and New York: Oxford Univer. Press, 1927.

Peters, H. N., & Murphree, O. D. The conditioned reflex activity in the chronic schizophrenic. *J. clin. Psychol.*, 1954, **10**, 126-130.

Pfaffman, C., and Schlosberg, H. The conditioned knee-jerk in psychotic and normal individuals. *J. Psychol.*, 1936, **1**, 201-205.

Prokasy, W. F., Hall, J. F., and Fawcett, J. T. Adaptation, sensitization, forward and backward conditioning, and pseudo-conditioning of the GSR. *Psychol. Rep.*, 1962, **10**, 103-106.

Razran, G. H. S. Conditioned responses: an experimental study and a theoretical analysis. *Arch. Psychol.*, 1935, No. 191.

Razran, G. H. S. Attitudinal control of human conditioning. *J. Psychol.*, 1936, **2**, 327-337.

Richter, C. P., and Wada, T. Method of measuring salivary secretion in human beings. *J. lab. & clin. Med.*, 1924, **9**, 271.

Rodnick, E. H. Characteristics of delayed and trace conditioned responses. *J. exp. Psychol.*, 1937, **20**, 409-425.

Rozhdestvenskaia, V. I., Nebylitsyn, V. D., Borisova, M. N., and Ermolaeva-Tomina, L. B. A comparative study of various indicators of the strength of the human nervous system. *Soviet Psychol. & Psychiat.*, 1963, **1**, 10-22.

Schiff, E., Dougan, C., and Welch, L. The conditioned PGR and the EEG as indicators of anxiety. *J. abnorm. soc. Psychol.*, 1949, **44**, 549-552.

Shipley, W. C. Studies of catatonia: VI. Further investigation of the perseverational tendency. *Psychiat. Quart.*, 1934, **8**, 736-744.

Silverman, R. E. Eliminating a conditioned GSR by reduction of experimental anxiety. *J. exp. Psychol.*, 1960, **59**, 122-125.

Sokolov, E. N. Neuronal models and the orienting reflex. In Mary A. B. Brazier (Ed.), *The central nervous system and behavior*. Madison, New Jersey: Madison Printing Co., 1960.

Solyom, L., and Beach, L. Further studies upon the effects of the administration of ribonucleic acid in aged patients suffering from memory (retention) failure. *Neuropsychopharmacol.*, 1961, **2**, 351-355.

Spence, K. W. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127-134.

Spence, K. W., and Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116-119.

Spence, K. W., and Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376-385.

Spence, K. W., and Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.

Spence, K. W., and Taylor, J. A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265-272.

Spence, K. W., Farber, I. E., and Taylor, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404-408.

Spooner, A., and Kellogg, W. N. The backward-conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321-334.

Stern, J. A., Das, K. C., Anderson, J. M., Biddy, R. L., and Surphlis, W. Conditioned alpha desynchronization. *Science*, 1961, **134**, 388-389.

Stewart, M. A., Winokur, G., Stern, J. A., Guze, S. B., Pfeiffer, E., and Hornung, F. Adaptation and conditioning of the galvanic skin response in psychiatric patients. *J. ment. Sci.*, 1959, **105**, 1102-1111.

Stewart, M. A., Stern, J. A., Winokur, G., and Fredman, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68**, 60-67.

Sweetbaum, H. A. Comparison of the effects of introversion-extraversion and anxiety on conditioning. *J. abnorm. soc. Psychol.*, 1963, **66**, 249-254.

Switzer, S. A. Anticipatory and inhibitory characteristics of delayed conditioned reactions. *J. exp. Psychol.*, 1934, **17**, 603-620.

Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.

Taylor, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303-320.

Taylor, J. A., and Spence, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497-502.

Teplov, B. M. Typological properties of the nervous system and their psychological manifestations. In N. O'Connor (Ed.), *Recent Soviet psychology*. New York: Pergamon Press, 1961.

Vogel, M. D. The relation of personality factors to GSR conditioning of alcoholics. An exploratory study. *Canad. J. Psychol.*, 1960, **14**, 275-280.

Vogel, M. D. GSR conditioning and personality factors in alcoholics and normals. *J. abnorm. soc. Psychol.*, 1962, **63**, 417-421.

Voronin, L. G., and Sokolov, E. N. Cortical mechanisms of the orienting reflex and its relation to the conditioned reflex. In H. H. Jasper & G. D. Smirnov (Eds.), Moscow Colloquium on Electroencephalography of Higher Nervous Activity. *EEG clin. Neurophysiol.*, 1960, Suppl. No. 13.

Welch, L., and Kubis, J. Conditioned PGR (psychogalvanic response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372-381.

Wells, C. E., and Wolff, H. G. Formation of temporary cerebral connections in normal and brain-damaged subjects. *Neurology*, 1960a, **10**, 335-340.

Wells, C. E., and Wolff, H. G. Electrographic evidence of impaired brain-function in chronically anxious patients. *Science*, 1960b, **131**, 1671-1672.

White, C. T., and Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357-362.

Wickens, D. D., and Wickens, C. D. Some factors related to pseudoconditioning. *J. exp. Psychol.*, 1942, **31**, 518-526.

Wickens, D. D., Allen, C. K., and Hill, F. A. Effect of instructions and UCS strength on extinction of the conditioned GSR. *J. exp. Psychol.*, 1963, **66**, 235-240.

Willett, R. A. Measures of learning and conditioning. In H. J. Eysenck (Ed.), *Experiments in personality*. London: Routledge & Kegan Paul, 1960, Vol. II.

Willet, R. A. Eyeblink conditioning and situation induced anxiety drive. In H. J. Eysenck (Ed.), *Experiments in motivation*. New York: Pergamon Press, 1963.

Wolpe, J. Experimental neuroses as learned behaviour. *Brit. J. Psychol.*, 1952, **43**, 243-268.

Wolpe, J. *Psychotherapy by reciprocal inhibition*. Stanford, California: Stanford University Press, 1958.

Wyrwicka, W., and Dobrzecka, C. Relationship between feeding and satiation centers of the hypothalamus. *Science*, 1960, **132**, 805-806.

Yates, A. J. The application of learning theory to the treatment of tics. *J. abnorm. soc. Psychol.*, 1958, **56**, 175-182.

REPRESSION-SENSITIZATION AS A DIMENSION OF PERSONALITY

Donn Byrne

THE UNIVERSITY OF TEXAS, AUSTIN, TEXAS

Psychologists in the field of personality have evidenced a consistent and long-continued interest in unconsciously motivated behavior. It may be observed that some individuals fail to verbalize feelings of anxiety when confronted by stimuli judged by outside observers to be anxiety evoking. Further, these same individuals are seen subsequently to engage in activities which may be hypothesized to control or reduce anxiety. Observations and speculations of this variety have stimulated a considerable amount of research, especially during the past two decades.

Anxiety-reducing activities which are unconsciously motivated are given the label of defense mechanisms. Various descriptions of these mechanisms, originating almost exclusively in psychoanalytic theory, have been proposed over the years. A unidimensional categorization which encompasses many diverse mechanisms grew out of the research of the new look in perception in the 1940's: repression-sensitization. At one end of this continuum of defensive behaviors are those responses which involve avoidance of the anxiety-arousing stimulus and its consequents. Included here are repression, denial, and many types of rationalization. At the sensitizing extreme of the continuum are behaviors which involve

an attempt to reduce anxiety by approaching or controlling the stimulus and its consequents. The latter mechanisms include intellectualization, obsessive behaviors, and ruminative worrying.

Beginning with his 1950 doctoral dissertation at Stanford, Charles Eriksen was one of the first investigators to make and to pursue this distinction of types of defense mechanisms while Jesse Gordon (1957) introduced the terms "repressor" and "sensitizer." The specific research which formed the background for much of the current work on the repression-sensitization dimension will be reviewed briefly.

I. Approach versus Avoidance Responses to Threatening Stimuli

A. PERCEPTUAL DEFENSE

1. Initial Investigations

An early and extremely influential article in the postwar perceptual studies was co-authored by Jerome Bruner and Leo Postman (1947a). The basic assumption underlying most of the new-look approach was stated at the outset of that article (p. 300):

"Perception is a form of adaptive behavior. Its operation reflects not only the characteristics of sensorineural processes, but also the dominant needs, attitudes, and values of the organism. For perception involves a *selection* by the organism of a relatively small fraction of the multiplicity of potential stimuli to which it is exposed at any moment in time. In perception, moreover, certain stimuli are *accentuated* and vivified at the expense of others. Finally, what is "habitually seen" in any given perceptual situation is a function of the *fixation* of past perceptual responses in similar situations. Through these three processes—selection, accentuation, and fixation—the adaptive needs of the organism find expression in perception."

Individual differences in perceptual adaptation to threat soon became evident in investigations emerging from the Laboratory of Social Relations at Harvard. With 19 undergraduates as subjects, Bruner and Postman (1947b) obtained associative reaction times for each of 99 words including a large proportion of potentially threatening ones such as "raped," "death," "penis," "agony," and "blush." Two weeks later, each subject was presented with 18 stimulus words (those yielding the individual's six fastest, midmost, and slowest reaction times) on a tachistoscope. Each word was presented at increasingly slower exposure speeds until correct recognition was obtained. A significant relationship between reaction time and recognition exposure speeds was found. Of primary

interest here was the finding of two different patterns of response among the subjects. In attempting to achieve an explanation of the two patterns, the authors suggested that some individuals revealed a defense process in which recognition threshold was a monotonic increasing function of associative reaction time. The greater the anxiety, the greater the "perceptual defense." The responses of other subjects suggested a sensitizing process in which recognition time was actually faster for the most anxiety-provoking words than for the six midmost words. Rather than avoiding the perception of threat, the latter individuals were characterized as vigilant in perceiving it. Perceptual vigilance had been found previously in response to goal-relevant stimuli congruent with need reduction (Bruner and Goodman, 1947) and value systems (Postman, Bruner, and McGinnies, 1948), but the notion of sensitization to threat created a new wave of research interest.

2. Individual Differences in Perceptual Defense

The bulk of the perceptual defense literature which followed the Harvard studies dealt with demonstrations, explanations, and criticisms of findings reporting threshold differences for threat versus nonthreat stimuli. In addition, however, individual differences in characteristic defensive reactions along a repressing-sensitizing dimension were also under investigation. In general, research was conducted in which subjects were categorized as to defensive mode by one means or another and then tested for perceptual threshold differences for threat versus nonthreat stimuli (Carpenter, Wiener, and Carpenter, 1956; Eriksen, 1951, 1952a; Kissin, Gottesfeld, and Dickes, 1957; Kogan, 1956; Kurland, 1954; Lazarus, Eriksen, and Fonda, 1951; Nelson, 1955; Perloe, 1960; Shannon, 1962).

The subjects in these studies consisted of the usual groups of college undergraduates, hospitalized neuropsychiatric patients, and ambulatory clinic patients. Differential recognition thresholds or differential accuracy scores were obtained for a variety of neutral and conflictful (primarily sex and hostility) stimuli which included words, drawings, Blacky pictures, and sentences recorded on tape. Recognition was made difficult by variations in speed of exposure, intensity of illumination, legibility of successive carbon copies, loudness of recordings, and intelligibility of auditory presentations partially masked by white noise. Independent indicators of defensive mode included readiness to verbalize conflictful material on sentence completion tests and in TAT stories, ability to recall failure versus success words in an experiment using scrambled sentences, ability to recall Blacky pictures, scores on the California F Scale, scores on the Defense Preference Inquiry for the Blacky pictures, clinical

ratings of Rorschach and Machover Figure-Drawing protocols, and classi-
fication as to characteristic defensive mode on the basis of interviews and
case history material.

Although these studies utilized a variety of subject populations,
perceptual tasks, and measures of defenses, significant relationships were
consistently reported between perceptual behavior and defenses with few
exceptions (e.g., Kurland, 1954). Those individuals who have difficulty
in perceiving threatening material accurately also give evidence of block-
ing, repression, and avoiding when responding to conflictful stimuli in
other contexts. Conversely, those who perceive threatening stimuli as
accurately as or more accurately than neutral stimuli respond in other
situations with intellectualization, sensitization, and general approach
behavior.

It should be noted that the explanation for these findings is far from
clear. As with other perceptual defense data, some investigators prefer to
explain the findings in terms of variables other than defense mechanisms
and unconscious motives. For example, after reviewing the literature on
individual differences, Brown (1961, p. 323) suggests:

"The individual-difference method does not, however, preclude the
possibility that the reactions to all tests are due to differential experiences
with the materials of the tests rather than to percepual defense. When
individuals who are "expressive" with respect to sexual matters are selected
out of a group, we may simply be selecting those who, in addition to
being expressive, or perhaps because of their expressiveness, have had
more extensive experiences with sexual words, symbols, and ideas. On
the other hand, the test-defined 'sexual inhibitor' might well have had
fewer exposures to, and experiences with materials of a sexual nature.
If such groups differ in their ability to identify sexual words, it might
be more sensible to explain these results by referring to individual dif-
ferences in associative strength rather than to traits of defensiveness or to
value systems with vague motivational overtones."

Given such an explanation, the mysterious quality of perceptual
defense vanishes in the clear air of the known relationship between word
familiarity and perceptual accuracy (e.g., Solomon and Postman, 1952).
Thus, theoretical speculations about the perceptual data have tended to
move away from some of the more colorful earlier notions such as un-
conscious perception. We are nevertheless, as Lazarus et al. (1951) point
out, left with individual differences in expressiveness, experiences with
conflictful material, defenses, or whatever label one chooses to apply. For
personality psychologists, the questions still remain as to the antecedents,
correlates, and consequents of these individual differences.

3. Repression-Sensitization in Other Response Measures

At approximately the same period in time as the differential thresh-old studies, many other investigations utilizing repressing-sensitizing variables were conducted. It was found, for example, that subjects who tend to recall their failures in an experimental task tend to recall material associated with a painful shock, while those who forget one forget the other (Lazarus and Longo, 1953). Individuals who recall incompleted tasks in a threatening situation learn affective words as easily as neutral ones, while those who forget the incompleted tasks experience relative difficulty in learning affective words (Eriksen, 1952b). Those subjects who are able to verbalize a pattern of electric shock applied during a learning task are able deliberately to avoid giving the punished responses (Eriksen and Kuethe, 1956). Relatively shorter latency for aggressive and succorant words on a word-association test is associated with the acceptance of such concepts on the Rorschach, while relatively longer latency is associated with failure to accept the concepts on the Rorschach (Eriksen and Lazarus, 1952). Giving emotional words in response to appropriate TAT cards is positively related to the number of Rorschach responses given (Ullman, 1958). Patients identified as facilitators according to their case histories prefer sexual and aggressive humor while patients identified as inhibitors prefer nonsense humor (Ullman and Lim, 1962).

An examination of the perceptual studies and the subsequent work suggests rather strongly the presence of an approach-avoidance sort of dimension with respect to response to threatening stimuli. It should also be noted that these behavior tendencies appear to be fairly pervasive ones in that they are identifiable in perceptual responses, responses given to projective tests, behavior in learning and memory tasks, and in symptoms of maladjustment. Such relational fertility is a convincing argument for the value of pursuing this variable in further research.

B. Measurement of Repression-Sensitization

A familiar problem which arises, though, is that of agreeing upon a suitable defining operation for repression-sensitization. In the work outlined above, almost every investigator utilized a different measure of the defense dimension. And, while significant correlations were generally reported between the various approach-avoidance indices, the magnitude of the relationships is hardly sufficient to conceptualize them as interchangeable measures of a single construct. An additional difficulty with the measures used in the various experiments is the problem of low or even unknown reliability (Byrne and Holcomb, 1962). As a first step in establishing the presence of a relationship between variables, relatively

low reliability of measurement only acts to increase the probability of making Type II errors. It is essential, however, in order for research in an area to progress beyond these first crude findings that Y is related to X, to obtain reliable measures of the variables involved and to utilize operations consistently across experiments.

If the preceding assumption is accepted, what would be the next logical step in measurement? One possibility would be to investigate the reliabilities of each reported defining operation (differential thresholds, memory scores, TAT responses, etc.) and select the most reliable as the measure of repression-sensitization in all future work. If satisfactory reliabilities were not found, the best measure could be used as a starting point for further test construction. The decision is a somewhat arbitrary one, but the goal in any event would be a consistent measuring instrument which yields scores that correlate with behaviors already defined as repressing-sensitizing. Consistency here means that the device has homogeneous response variations across items, stability of scores over time, and an objective scoring system on which any two independent judges can obtain close agreement. A measurement approach meeting those criteria is one based on a combination of repression-sensitization indices which have not yet been discussed.

1. Use of the MMPI to Measure Defensive Behavior

While the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway and McKinley, 1951) was originally devised as a diagnostic aid, it has subsequently been used in both research and clinical work for a wide variety of additional purposes. Since the individuals classified into the various psychiatric syndromes identified by the test tend to differ in characteristic defensive modes and since some of the validity scales have been considered as reflecting defenses, it was perhaps inevitable that MMPI scales would be utilized in work on repression-sensitization.

A summary of the scales and combinations of scales used by various investigators is shown in Table I. There was relatively good agreement among investigators as to the type of MMPI items most indicative of repressing and sensitizing defenses.

In the work cited in Table I, the MMPI measures of repression-sensitization were found to be related to such variables as differential recall of completed and incompleted tasks, frequency of dream recall, defense mechanisms in hospitalized neurotics, differential perceptual recognition thresholds, and interpersonal perception. With these findings as a promising start, two independent efforts were made at approximately

the same time to devise a single, comprehensive MMPI scale to measure the repression-sensitization dimension.

TABLE I
MMPI SCALES USED AS DEFENSE MEASURES

Scales	Represser score	Sensitizer score	Reference
K	High	Low	Page and Markowitz (1955)
L	High	Low	Page and Markowitz (1955); Tart (1962)
F minus K	Low	High	Ullmann (1958)
Hy	High	Low	Eriksen (1954); Mathews and Wertheimer (1958)
Hy denial	High	Low	Carlson (1954); Gordon (1957); Gordon (1959)
Hy admission	Low	High	Gordon (1959)
Hy minus Pt	High	Low	Eriksen and Davids (1955); Truax (1957)
Pt	Low	High	Carlson (1954); Eriksen (1954); Eriksen and Browne (1956); Eriksen and Davids (1955); Eriksen, Keuthe and Sullivan (1958)
MAS	Low	High	Eriksen and Davids (1955); Gordon (1959)
Welsh A	Low	High	Tart (1962)
Welsh R	High	Low	Tart (1962)

2. A Repression-Sensitization Scale

At the Behavioral Research Laboratory of the Veterans Administration Hospital at Palo Alto, a number of investigations were undertaken to follow up and extend Shannon's (1955, 1962) work with facilitators and inhibitors. Ullmann (1958) devised a scoring system for case history material which was used in several projects. As a primary measure, however, the case history scoring approach was limited by problems of interjudge reliability, the use of discrete categories of individuals rather than a continuum, and its inapplicability to nonpatient populations. To overcome these difficulties, Ullmann (1962) undertook the construction of an empirically derived MMPI scale to measure facilitation-inhibition (or repression-sensitization). Using an item-analysis approach, he compared the responses of 38 facilitators and 24 inhibitors (identified by the case history method) on each of the 566 MMPI items. A cross-validation sample consisted of 48 facilitators and 22 inhibitors. A total of 21 items met the criterion of differentiation at the .05 level in both samples. An additional 23

items differentiated at the .10 level *and* were related to the total facilitation-inhibition score in a third sample of 61 subjects. The final scale consisted of 44 items. On a new sample of subjects, a corrected split-half reliability of .96 and test-retest reliabilities of .88 (1 to 18 months), .71 (19 to 36 months), and .54 (37 to 85 months) were obtained. Studies utilizing Ullmann's scale will be reported in a following section.

At about the same time that Ullmann's work was under way, Altrocchi, Parsons, and Dickoff (1960) at Duke University approached the measurement of repression-sensitization via the MMPI in a somewhat different way. On the basis of the work of Eriksen and Gordon, three MMPI scales were selected as likely measures of sensitization (D, Pt, and Welsh Anxiety) and three others as measures of repression (L, K, and Hy denial). A repression-sensitization index was determined by subtracting each subject's total score on the latter three scales from his total score on the three former ones. In neither this paper nor a subsequent report (Altrocchi, 1961) were reliability data presented.

In work at the University of Texas, the author (Byrne, 1961, p. 337) attempted to refine Altrocchi's scale because "Several potential measurement difficulties arise with this measure because of item overlap among the six MMPI scales which are combined in it. For example, item 32 contributes to D, Pt, and Welsh Anxiety scores, thereby giving it an arbitrary weight of three. More confusing possibilities also arise. For example, item 30 contributes to D and to L, K, and Hy denial; thus, it is included in opposing halves of the index with a net weight of two for repression. While such differential item weights could conceivably prove to be the optimum ones, their accidental nature in this instance is clear."

To get around such difficulties, the author substituted a scoring system in which each item on the six scales was scored only once and all inconsistently scored items eliminated. The result was a 156-item Repression-Sensitization (R-S) Scale in which high scores indicated sensitizing responses. Both the corrected coefficient of internal consistency and the coefficient of stability (six weeks) were found to be .88. In subsequent work (Byrne, Barry, and Nelson, 1963), an internal-consistency item analysis was undertaken with two independent samples of 370 students each. Those 127 items which yielded correlations with the total R-S score significant at the .001 level or better in *both* samples were retained as scorable items in the revised R-S scale. This revised scale was found to have a corrected split-half reliability of .94 and a test-retest reliability (3 months) of .82.

The following section surveys the published and unpublished research in which these MMPI repression-sensitization measures have been used. Approximately 97% of Altrocchi's items were retained in the

original and about 70% in the revised R-S scale. The Facilitation-Inhibition Scale has been found to correlate .76 with the R-S scale in one student population (Byrne, 1961), .88 in another (Liberty, Lunneborg, and Atkinson, in press), and .94 in a neuropsychiatric patient population (Ullmann, 1962). Almost half of the items in the F-I scale also are contained in both the original and revised R-S scales. Therefore, it seems defensible to consider the instruments as interchangeable measuring devices.

II. A Review of Research on the R-S Scale

A. REPRESSION

Basic to the construct validity of the R-S scale is evidence relating scores on the test to repressive *versus* sensitizing behavior. As with most variables in the field of personality, the conceptualizations regarding defensive modes range from nonoperational theoretical formulations based on clinical observations to quite specific operations devised for research purposes. Since no single variable of the latter variety can encompass all of the surplus meaning contained in the clinical theories of repression, the construct validity of the R-S scale must rest on a series of correlational and experimental findings. Seven different types of evidence have been reported.

1. Judgments Concerning Defense Mechanisms

Dr. Vincent Tempone compared the R-S scale with judgments by clinicians of the meaning of this variable. He had nine staff members at the Massachusetts Mental Health Center fill out the test as they thought a repressing individual would. He defined agreement as concurrence by at least seven of the nine judges. On the original R-S scale, judges corresponded with the key on 72% of the items, and on the revised scale there was correspondence on 90%. On none of the items did the judges agree on a response opposite to the key (Byrne *et al.*, 1963).

A second investigation by Tempone (1963) utilized ratings of 18 therapy patients by six first-year psychiatric residents. Patients' scores on the R-S scale and clinical ratings of their repression-sensitization were found to correlate .38, which falls short of statistical significance in this small sample.

Davison (1963) administered several measures of coping mechanisms, including the R-S scale, to 48 male students. A cluster analysis of the 32 variables led to the identification of three clusters. The first was essentially defined by the R-S scale, and it included (in addition to the MMPI scales used in constructing the R-S scale) the *F, Pd,* and *Ma* scales from

the MMPI, intellectualization as judged in an interview situation, and Identification on the Heath Phrase Association Test.

2. Selective Forgetting

At the University of Arkansas, Gossett (1964) attempted to elicit repressive mechanisms by combining threatening conditions with a memory task. The subjects consisted of 48 repressers and 48 sensitizers drawn from a pool of 296 introductory psychology students. Each subject learned a list of 12 nonsense syllables. Their performance on the fifth trial constituted the first recall test. Following this, they were given an "Intelligence and Personality Test" made up of 12 subtests, each titled with one of the nonsense syllables. In the experimental group, failure was induced via false norms, and the failures were attributed to certain intellectual and/or personality problems. The control group took the test as a "standardization sample" without false norms or induced failure. After a filler task, all subjects were asked to recall the 12 syllables. As a control for suppression, half of the subjects were offered rewards for remembering correctly ($10 and the chance to avoid several hours of tedious tasks) while half were given standard instructions to attempt to remember the syllables. After an explanation of the purpose of the experiment to the subjects, a third recall test was given. Gossett found significant differences between repressers and sensitizers in their recall of the syllables when the effects of suppression were controlled. He interpreted this finding to mean that those with low scores on the R-S scale repress threatening material while those with high scores do not.

McReynolds and Ullmann (1964) with 40 students in an advanced psychology class obtained recall data for 36 words projected briefly on a screen immediately after the last word had been presented. Of the 30 nonpractice words, 12 were pleasant, 12 were unpleasant, and 6 were neutral. The recall score consisted of the difference between the number of unpleasant and pleasant words recalled. With anxiety, number of neutral words recalled, and general tendency to use pleasant or unpleasant words held constant, the correlation between scores on the F-I scale and the recall score was not significant.

3. Perceptual Defense

Considering the origin of the conceptualizations of repression-sensitization, it is obvious that scores on the R-S scale would be expected to be related to differential perceptual recognition thresholds for threatening *versus* nonthreatening stimuli. Tempone (1962) tested this proposition, utilizing an induced failure manipulation as the source of threat. From a pool of 244 University of Texas undergraduates, he selected 40

repressers (below 25th percentile) and 40 sensitizers (above 75th percentile) on the basis of scores on the R-S scale. On an eight-item anagrams task taken with a group of stooges, half of the subjects were given a failure experience (two or less anagrams correct) and half a success experience (six or more anagrams correct). The correct solution to each anagram was given to the subjects after each trial. Following this phase of the experiment, a perceptual test was given in which 19 words were presented at gradually increasing exposure speeds tachistoscopically; the stimuli consisted of three practice words, the eight correct anagram solu-
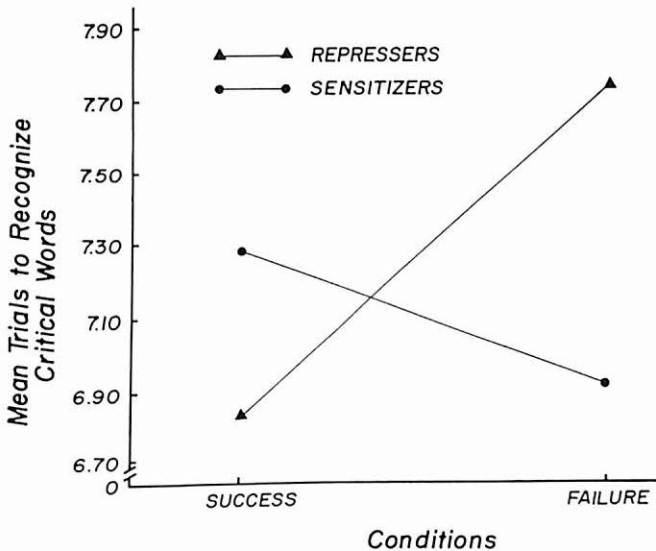
Fig. 1. Perceptual recognition thresholds for critical words under success and failure conditions. [After Tempone (1962).]

tions, and eight neutral words. The order of the latter 16 words was counterbalanced, and the critical and neutral words were matched for word length and frequency based on the Thorndike-Lorge word count. Word thresholds were measured as number of trials required for correct recognition. Analysis of variance indicated a significant interaction effect for repression-sensitization times success-failure. Repressers and sensitizers did not differ significantly in the success condition, but repressers obtained significantly higher thresholds for the threat words in the failure condition than did sensitizers. For the neutral words, thresholds were not related to defensive mode or experimental conditions. The relationship is shown graphically in Fig. 1.

Ullmann, Weiss, and Krasner (1963) were able to change the perceptual performance of repressers by means of a process of verbal condi-

tioning for use of emotional words just prior to responding to a perceptual defense task. The task was the recognition of matched neutral and threatening words on a series of successively better carbon copies. Compared to the control group, inhibitors in the experimental group showed significantly less difference in their recognition thresholds for threatening and neutral words; facilitators in the two groups did not show a significant difference. The change evidenced by those with repressive defenses could be interpreted as a function of increased familiarity with the threatening words, desensitization to the threat represented by such words, or as learning that the use of such words was socially acceptable in the context of the experiment.

4. Awareness of Anxiety

In his dissertation research at the University of Rochester, Pomeranz (1963) investigated the relationship between verbalized response to threat and repression-sensitization. Three types of movies (control, stressor, and alleviation of emotional arousal) were shown, and the subjects' affective states during each were obtained on an Adjective Check List and an Anxiety Differential measure. The subjects were 63 male undergraduates selected from a larger group on the basis of R-S scale scores. It was found that sensitizers indicated significantly more emotional arousal in response to the stressor (movie of unpleasant surgical operation) than did repressers.

Byrne and Sheffield (in press) created a potential threat situation by exposing subjects to 11 vividly descriptive sexual passages from various novels. Subjects were all males with 22 sensitizers and 22 repressers serving in the sex arousal condition. In the control group 22 sensitizers and 22 repressers read neutral descriptive passages from the same 11 novels. Immediately after the reading task, all of the subjects were asked to fill out six five-point rating scales dealing with their feelings while reading the literary material. On the "Sexually Aroused" scale, ratings were significantly higher in the experimental than in the control group for both repressers and sensitizers. On the "Anxious" scale, however, there was a significant interaction in that repressers and sensitizers did not differ in the neutral condition, but sensitizers rated themselves as significantly more anxious in the experimental condition. The relationship is shown in Fig. 2. A possible interpretation of this finding is that repressers were unaware of or were denying any feelings of anxiety as a concomitant of sexual excitement. It would be important, therefore, to be able to find some other evidence of disturbance accompanying the denial of anxiety. With experimental and control groups combined, self-ratings of sex arousal correlated with self-ratings of disgust ($r = .34$, $p < .05$) and

anger ($r = .36$, $p < .05$) for repressers. For sensitizers, sex arousal was significantly related to anxiety ($r = .67$, $p < .01$), lack of boredom ($r = -.57$, $p < .01$), and feeling entertained ($r = .33$, $p < .05$).

Relevant data are also presented by Lomont (1964) in an investigation utilizing 35 hospitalized psychotic and nonpsychotic patients. During their first month of hospitalization, the subjects were given the R-S scale, a 66-item word association test containing both threatening and nonthreatening words, and the IPAT Self Analysis Form. The word association test was scored for 31 signs of disturbance including reaction
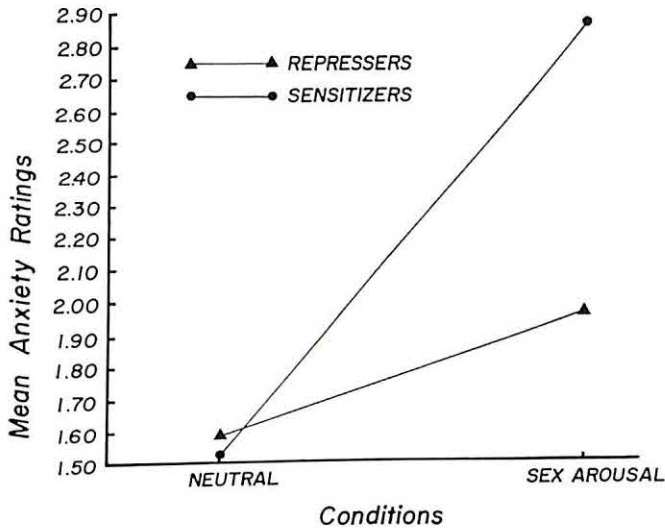
FIG. 2. Self-ratings of anxiety after reading neutral and sexually arousing literary material. [From Byrne and Sheffield (in press).]

time over 2.5 seconds, blocking, and reproduction failures. A significant positive correlation was found between R-S scale scores and IPAT scores ($r = .76$) and a significant negative correlation between the R-S scale and disturbance scores on the word association test ($r = -.45$). Thus, sensitizers responded with greater verbalized anxiety than repressers on a self-report measure while the repressers indicated greater disturbance than sensitizers on the indirect measure.

5. Physiological Response to Threat

Lazarus and Alfert (in press) report an investigation in which cognitive appraisal of a stress situation and physiological response to that stress were compared. In previous research (Lazarus, Speisman, Mordkoff, and Davison, in press; Speisman, Lazarus, Mordkoff, and Davison, in press), a silent movie of a primitive subincision ritual was used to produce

stress reactions. In the Lazarus and Alfert study, three experimental conditions were employed: the silent film without introduction or commentary, the film with both an introduction and a movie sound track that emphasized denial and reaction formation, and the silent film with the denial and reaction formation introduction. The latter two conditions were conceptualized as possibly alternate ways to reduce the threat of the film; there was denial of the painful aspects of the operation and of its possible physical harm and an emphasis on the joy of the native boys in participating in the ceremony. The subjects were 69 male students at the University of California. During the movie, measurements of skin conductance and heart rate were made continuously. Immediately after the film, several psychological measures were obtained including the Nowlis Adjective Check List of Mood (Nowlis and Nowlis, 1956). In a separate research session, the subjects were given the MMPI which was scored for K, Hy denial, Welsh Repression, and Repression-Sensitization. In general they found on both the physiological measures and mood ratings that the silent film condition was the most stressful and the silent film plus denial introduction the least stressful. Scores on the four personality variables were divided at the median. In the various comparisons which were made, the K and Hy denial scales were the most effective predictors of individual differences. The R-S scale did not yield as many significant findings (possibly because of a median split rather than the use of extreme groups), but it followed the same general pattern as the other variables. The over-all tendency was for those individuals with denying and repressing defenses to show higher levels of skin conductance and lower levels of discomfort on the verbal measures when compared with subjects with sensitizing defenses. The skin conductance data for the R-S scale are shown in Fig. 3. Even though Lazarus and Alfert found that those with low scores on the R-S scale indicated less anxiety and depression on the Nowlis Mood Scale than did those with high scores, the reverse pattern was found for skin conductance. The authors conclude:

"Judging from these Nowlis patterns, verbally derived measures of stress response in the form of disphoric affect interacts with personality variables in a direction opposite from what is found with autonomic indicators. High deniers refuse to admit disturbance verbally but reveal it autonomically, while low deniers are apt to say they are more disturbed while showing less autonomic reactivity."

In his doctoral research at the University of California, Davison (1963) also utilized the subincision ritual movie as a stressor. A group of 48 male undergraduates viewed the film weekly for a total of three

exposures. On each occasion, recordings of several physiological responses were obtained. Scores on the repression-sensitization dimension predicted reaction to the film in terms of skin conductance, heart rate, and bodily movement. Dividing the subjects into high, medium, and low groups, he found that the medium group was the most physiologically responsive to the film while the sensitizer group was the least responsive. A possible implication is that the development of strong consistent defenses of either type protects the individual from physiological stress, but that sensitization is more effective in this respect than repression. The
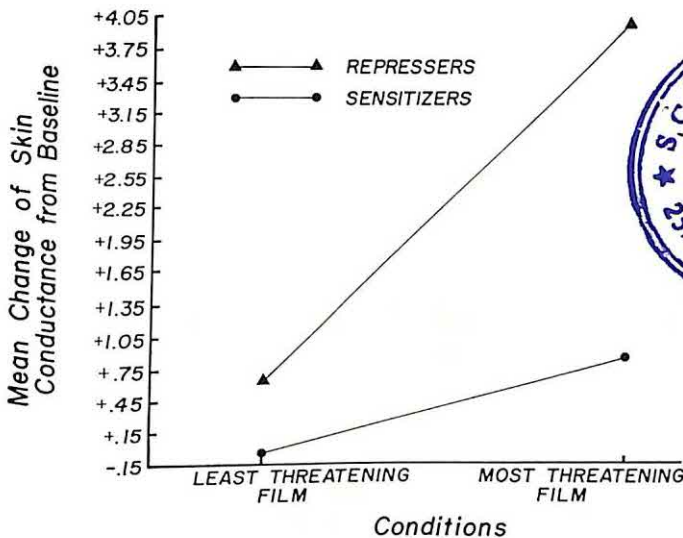
FIG. 3. Changes in skin conductance from baseline to least threatening and most threatening film conditions. [From Lazarus and Alfert (in press).]

greatest differences were between the sensitizer group and the other two; Davison (1963, p. 123) suggests,

"We have some support for the construct validity of the Byrne scale as a measure of sensitization in two facts: (1) the high sensitizers show least movement during the film and our previous findings associate immobility in this situation with attention; (2) the high sensitizers obtain the largest first-session mean on Nowlis "Activation," which on its content would seem to reflect self-reported mental alertness and interest, and they demonstrate a progressive decline in this dimension. In other words, they appear to be reacting to the first exposure to the film, a threatening situation, with mobilization of their preferred coping mechanisms, and this response diminishes as the film loses its threatening capability."

6. *Responding to Ambiguous Stimuli*

Responses to relatively ambiguous or unstructured stimuli have been found to be in part a function of the subjects' characteristic defense mechanisms. Those who repress tend to respond to ambiguous material with neutral or nonthreatening interpretations, while individuals employing intellectualizing or obsessional defenses respond to such material with conflict-laden and emotional content. Are scores on the R-S scale related to these differences? If repression-sensitization scores indicate a general tendency to avoid or approach threatening stimuli and if sexual and aggressive cues represent some degree of threat to almost everyone in our culture, interpretation of projective material as sexual or aggressive should be a function of R-S scale scores.

From a group of 213 undergraduates, 29 sensitizers and 24 repressers were asked to respond to nine TAT cards (Byrne, 1961). Each protocol was scored for sexual content, aggressive content, and percentage of emotional words. Neither the aggression score nor the percentage of emotional words was related to repression-sensitization. For both sexes, sensitizers had higher sexual scores than repressers. The difference for males was significant at beyond the .01 level. The females had a larger mean difference than the males, but with a smaller N and a larger variance the difference did not reach statistical significance.

A TAT investigation by Tempone (1963) yielded quite different results. With 46 male undergraduates at the University of New Hampshire as subjects, stories given in response to five TAT cards were scored for aggressive and sexual content. In this study, sensitizers gave significantly ($p < .05$) more aggressive content than repressers, but the differences in sexual responses were not significant.

In a word association study, Blaylock (1963) utilized a 100-item list made up chiefly of homonyms with aggressive and neutral meanings. A group of 37 Texas undergraduates was given this test and the R-S scale. Repression-sensitization was found to correlate .37 ($p < .05$) with number of stimulus words perceived as aggressive. When this investigation was repeated with a similar word association test administered individually, no relationship to R-S scale scores was found in a group of 54 undergraduates.

7. *Response to Humor*

On the basis of Freudian theory, O'Connell and Peterson (1964) hypothesized differential response to humorous material as a function of repressing *versus* sensitizing defenses. Specifically, they predicted that individuals with low appreciation for humor are rigid repressers, those

with high appreciation for humor are normals (middle part of the R-S scale), and that sensitizers should fall in between the other two groups in humor appreciation. The humor test was composed of 17 jokes and anecdotes which had been unanimously classified by six clinical psychologists as falling in the category of humor (rather than aggressive or hostile wit or nonsense wit). In defining these categories for the judges, the authors indicated that ". . . humor is characterized by an *objectively* stressful situation in which the principal reacts with a jest which is relatively free of hostility directed toward others." For example, one of the humor items was, "A condemned prisoner was being led to the gallows early one Monday morning. As he left his cell, he waved to the other prisoners and said, 'Well, this is a good beginning to the week.'" In its final form, the humor test required the subjects to respond to each item on a five-point scale ranging from "dislike very much" to "like very much"; its corrected split-half reliability was found to be .96.

TABLE II

REPRESSION-SENSITIZATION AND HUMOR APPRECIATION[a]

| Humor appreciation | N | R-S scale | |
		M	SD
High	54	39.20	17.82
Medium	70	42.20	18.36
Low	54	36.10	17.42

[a] After O'Connell and Peterson (1964).

Both the humor test and the R-S scale were administered to 178 students at Baylor University. The sample was divided approximately into thirds on the basis of their humor appreciation scores. As shown in Table II, the mean repression-sensitization scores for these three groups fell in the predicted order. A simple analysis of variance did not yield a significant F, but group comparisons by means of t tests indicated the low-appreciation group was significantly more repressing than either the middle- or high-appreciation groups.

8. Conclusions

Predictions based on the assumption that scores on the R-S scale indicate individual differences in the tendency to repress or deny or avoid threatening stimuli have been relatively well supported. There is evidence that behavior which clinicians define as repressive is related to the behavior measured by the test. When confronted by threatening or anxiety-provoking situations, individuals on the two ends of the scale differ in the predicted direction in terms of memory, perception, and reported anxiety; further, these differences are not manifest in neutral

situations. It has not yet been conclusively demonstrated that repressers give physiological evidence of anxiety while verbally denying such feelings, but the work of Lazarus and others offers some encouragement along this line. Response to projective material with emotionally disturbing content is not consistently related to repression-sensitization. It is possible that the inconsistent results in the latter work may be a function of the degree of threat provided by the stimuli. It would be hypothesized, therefore, that if the content of the stimulus material and/or the accompanying instructions and experimental setting were sufficiently threatening, differences in repression-sensitization would lead to differences in response. Finally, the relative dislike of humor by repressers as compared with other individuals is consonant with Freudian humor theory. In their excellent review of the general concept of repression, MacKinnon and Dukes (1962) describe three experimental approaches to the study of this behavior. Within that categorization system, the R-S scale has been found relevant in predicting (1) inhibition of perception, (2) inhibition of memory, and (3) inhibition of response. The construct validity of the scale as a measure of differences in defenses seems moderately well established.

B. MEASURES OF SELF-DESCRIPTION

On several bases it has been hypothesized that repression-sensitization is related to measures of the self concept. Individuals on the repressing end of the continuum would be expected to be unaware of those aspects of themselves which involve anxiety-arousing cues. Thus, in response to the typical assessment devices utilized in research on self theory, repressers should be less likely than sensitizers to indicate negative self attributes or to indicate discrepancies between their actual and ideal attributes. Sensitizers, in contrast, would be expected not only to be aware of negative attributes and self-ideal discrepancies but also to dwell on them and even to exaggerate them.

1. Self-Ideal Discrepancy

A number of earlier investigations reported that repressing individuals tend to describe themselves more positively on Q-sort tasks than do sensitizing individuals (e.g., Block and Thomas, 1955; Chodorkoff, 1954). With the R-S scale, the self-ideal discrepency of repressers on Leary's Interpersonal Check List was found to be significantly less than that of sensitizers in two different samples of student nurses (Altrocchi et al., 1960). The magnitude of the relationship is a relatively substantial one; on Worchel's (1957) Self-Activity Inventory (SAI) the R-S scale has been found to correlate with self-ideal discrepancy scores .62 ($p < .01$) and

.55 ($p < .01$) in two groups of undergraduates (Byrne, 1961) and .63 ($p < .01$) in a third sample (Byrne *et al.*, 1963). The self-ideal discrepancy scores on Bill's Index of Adjustment and Values and the R-S scale were found to correlate .61 ($p < .01$) in a group of 227 nursing students in New Jersey (Hanson, 1963).

2. Self Concept

Presumably, self-ideal discrepancy can come about as a function of variations in either self concept or ideal concept. It is reasonable to inquire, then, whether represser-sensitizer differences in discrepancy are a function of differences in descriptions of self, ideal, or both. Altrocchi *et al.* (1960) reported that repressers and sensitizers did not differ in ideal self, only in self concept. Sensitizers were found to have a more negative self concept and hence a greater discrepancy between self and ideal. Sensitizers described themselves as more rebellious, aggressive, and self-effacing, and as less dominant than repressers.

These findings were verified in subsequent research using the SAI. Each item on the latter was rated in terms of whether or not the behavior represented a negatively valued characteristic. On this basis, both self-descriptions and ideal-self descriptions could be scored in terms of negativeness. In two different investigations (Byrne, 1961; Byrne *et al.*, 1963), repression-sensitization was found not to be significantly related to negative ideal-self descriptions ($r = .25$; $r = .16$) but substantially related to negative self-descriptions ($r = .66$, $p < .01$; $r = .68$, $p < .01$). It seems clear that the findings regarding repression-sensitization and self-ideal discrepancy are a function of differences in self-descriptions. Repressers and sensitizers do not differ systematically in their ideals. Rather, it is their self-pictures which differ markedly, with repressers presenting themselves positively and sensitizers negatively.

3. Incongruency

A somewhat different sort of self-descriptive measure is that of incongruency or inconsistency of percepts. As has been discussed elsewhere (Byrne *et al.*, 1963, p. 330),

"McReynolds (1956) has proposed that anxiety is a function of the quantity of perceptual material not assimilated into perceptual systems. He uses the term "incongruency" to describe unassimilated material. Incongruent percepts are those which are difficult to harmonize because they are contradictory, inconsistent, or dissonant (McReynolds, 1958), and the existence of such percepts motivates individuals to reduce incongruencies. In a slightly different terminology, a stimulus is incongruent for an individual if it elicits incompatible responses. The concept of

incongruency is quite similar to Festinger's (1957) formulation of cognitive dissonance."

Incongruency has been operationally defined as inconsistency in self-report between values (good-bad) and feelings (like-dislike) with respect to various types of behavior, events, or objects. In order to measure incongruency, several researchers have constructed tests consisting of a series of items which take the form "Hunting for game with a gun" and "Insisting on my rights." Ss respond to the items in terms of values by rating each activity as good or bad. Later, they respond to the same items in terms of feelings by rating each as something that they like or dislike. Each item is then scored for congruency-incongruency across the two sets of ratings. A congruent response (Good-Like or Bad-Dislike) receives no points. An incongruent response (Good-Dislike or Bad-Like) receives one point. The total incongruency score is the number of items to which S has responded in an incongruent way. Among other findings, incongruency has been found to be positively related to scores on the Manifest Anxiety Scale (McReynolds, 1958). Also, incongruency with respect to hostility was found to be inversely related to the accuracy of perception of hostile cartoons while sexual incongruency was found to be inversely related to the accuracy with which sexual cartoons are identified (Byrne, Terrill, and McReynolds, 1961).

Since incongruency is conceptualized as anxiety arousing, it would follow that repressers should be less aware of their incongruent percepts than are sensitizers. In extending the incongruency concept, three types of incongruency were investigated. A series of 50 hostility items were used, and subjects rated each on the basis of their feelings of like-dislike, their values of good-bad, and their estimation of environmental consequences as pleasant-unpleasant. Thus, three types of incongruency were possible: feelings versus values (like-bad, dislike-good), values versus consequences (good-unpleasant, bad-pleasant), and feelings versus consequences (like-unpleasant, dislike-pleasant). A group of 114 undergraduates was given the R-S scale and the Hostility Incongruency Test. Each type of incongruency was significantly related to repression-sensitization: .38 ($p < .01$) for feelings versus values, .33 ($p < .01$) for feelings versus consequences, and .19 ($p < .05$) for values versus consequences. Further, in each of the former two instances, one of the two possible types of incongruency entirely accounted for the relationships. R-S scale scores correlated with number of items checked like-bad ($r = .40$, $p < .01$) but not with items checked dislike-good ($r = .00$), and repression-sensitization correlated with number of items checked like-unpleasant ($r = .45$, $p < .01$) but not with items checked dislike-pleasant ($r = -.03$). Sensitizers, then, are more likely to indicate that they enjoy behavior which they believe to

be morally wrong and which they believe is likely to have unpleasant environmental consequences.

4. Hostility

Using the Rosenzweig P-F test and a 14-item rating scale as hostility measures, Altrocchi, Shrauger, and McLeod (1964) found that repressers attributed less hostility to themselves than did sensitizers or expressers. In research which is as yet unpublished, Lomont (1963) investigated the relationships among the R-S scale, a self-report measure of hostility (the Buss–Durkee Inventory), and hostility as measured by the Holtzman Ink Blot Test. With 27 college students, repression-sensitization correlated .74 ($p < .01$) with scores on the Buss–Durkee scale but only negligibly with hostility as measured by the ink blots. A more unusual finding, however, had to do with the fact that repression-sensitization acts to obscure the relationship between the two hostility variables. The latter do not correlate significantly ($r = .28$) unless scores on the R-S scale are partialled out; in the latter case, the two hostility measures correlated .49 ($p < .05$). Lomont concludes:

"These results, I think, are nicely consonant with the clinical assumption that an ink blot test can tap unacknowledged hostility. In keeping with the implications of the clinical belief, repression appears to be an intervening variable which affects the correspondence between self-reports of aggression and aggression measures on an ink blot test. This picture of repression as an intervening variable implies that, given a certain degree of self-reported aggression, the degree of hostility which shows up on an ink blot test is positively related to the degree of repression operating on the self-reported aggression. . . . The results fit the clinical belief regarding the relation of repression to self-reported and ink blot aggression particularly well since R-S is highly related to self-reported aggression but not at all with ink blot aggression. This finding is in keeping with the clinical hypothesis that repression censors an individual's self-report of aggression but that the ink blot measure of aggression largely circumvents censorship."

5. Conclusions

In the research conducted to date, the relationship between R-S scale scores and measures of self-description is clearly and firmly established. The higher an individual's score on the R-S scale, the more likely he is on other instruments to indicate a discrepancy between his self concept and his self ideal, to describe himself in negative and self-depreciating terms, to report incongruencies among feelings, values, and environmental consequences, and more specifically to say that he enjoys

behavior which he believes to be morally wrong and which is likely to have unpleasant consequences. Further, as R-S scale scores increase, the more likely an individual is to describe his own behavior and feelings as hostile. These findings may represent differences in the tendency to present oneself in the most favorable or most unfavorable possible light.

C. MALADJUSTMENT

If each end of the repression-sensitization continuum represents an extreme of the respective defensive modes, scores on the R-S scale would be expected to have a curvilinear relationship with various indices of psychological adjustment. Neither overintellectualization of conflicts nor denial of them should result in optimal adjustment.

1. Population Comparisons

Because of the multitude of differences involved besides that of mal-adjustment, comparison of hospitalized neuropsychiatric patients with college students provides equivocal evidence at best. Nevertheless, the

TABLE III

REPRESSION-SENSITIZATION IN ALCOHOLICS AND NORMALS[a]

	Normals			Alcoholics			Mean difference	CR
	N	M	SD	N	M	SD		
Males	394	63.08	17.71	69	75.58	16.77	12.50	6.19[b]
Females	230	61.80	16.20	11	77.82	17.69	16.22	2.99[c]

[a] Data from Gynther (1963).
[b] $p < .001$.
[c] $p < .01$.

curvilinear hypothesis leads to the prediction that neuropsychiatric patients should have more extreme scores on the repression-sensitization measure and hence a larger standard deviation than normal subjects. Ullmann (1962) compared scores on the F-I scale for 47 male Texas students ($M = 29.38$, $SD = 6.50$) with those of two samples ($N = 90$ and 64) of male neuropsychiatric patients at the Veterans Administration Hospital at Palo Alto ($M = 25.74$ and 25.39, $SD = 11.22$ and 11.44). The patients had a significantly higher standard deviation than the students, as predicted. In addition, mean differences were also significant; the patients were further toward the facilitating or sensitizing end of the continuum.

In a group of 35 undergraduates, Joy (1963b) found a correlation of .91 between scores on the R-S scale and scores on an MMPI scale which measures Alcoholic Tendency. Because of item overlap on the two in-

struments, the meaning of this finding is open to question. Confirmation of the relationship between sensitizing defenses and alcoholism has been provided, however, by Gynther (1963) at the Malcolm Bliss Mental Health Center in St. Louis. As part of an investigation of married and unmarried alcoholics which is now in progress, Dr. Gynther obtained R-S scale scores on 69 male and 11 female alcoholics. As shown in Table III, for both sexes the alcoholics are significantly more sensitizing on the original R-S scoring system than the college students in the normative sample.

2. Deviation from the Norm

In his work on the Deviation Hypothesis, Berg (1959) has presented evidence to suggest that responses to test items which differ from those of the total group are indicative of abnormality. He has stated the Deviation Hypothesis (Berg, 1957, p. 159) as follows:

"Deviant response patterns tend to be general; hence those deviant behavior patterns which are significant for abnormality (atypicalness) and thus regarded as symptoms (earmarks or signs) are associated with other deviant response patterns which are in noncritical areas of behavior and which are not regarded as symptoms of personality aberration (nor as indicators, signs, earmarks)."

The tendency to respond deviantly, then, should be more characteristic of repressers and sensitizers than of those falling in the middle of this dimension. An adjective check list to measure deviancy was devised by Grigg and Thorpe (1960) by determining the 33 most commonly and 39 least commonly checked adjectives on Gough's Adjective Check List. Deviant scores consist of the number of adjectives of the first type *not* checked and the number of the latter that *are* checked. With two independent samples of undergraduates who had taken the R-S scale, deviant response scores on the Grigg–Thorpe list were obtained (Byrne, 1961). For one group the correlation was .42 ($p < .01$) and for the other .33 ($p < .01$) with no evidence of curvilinearity. Sensitizers tended to make deviant responses while repressers tended to make modal responses. It may be observed, though, that the commonly checked adjectives are primarily positive attributes and the least commonly checked adjectives negative attributes. If the adjective check list were viewed as simply a self-description measure, an alternative interpretation of the findings is that they represent an additional instance of sensitizers and repressers differing in the extent to which they describe themselves in positive or negative terms.

Further evidence on this point has been reported by Lucky and

Grigg (1964). Two measures of deviant response tendency were utilized, one which involved self-description (the Grigg–Thorpe adjective check list) and one which did not. The latter consisted of an "ESP" task in which subjects indicated for each of 72 items whether they believed that the experimenter wanted them to mark it; deviancy scores were defined as the number of items that deviated 1.5 SD's from the mean number of items circled by the group. Again, a positive linear relationship was found between R-S scale scores and deviant response scores on the adjective check list ($r = .40$, $p < .01$). With the ESP task, however, no relationship to deviancy was found ($r = .07$). It would appear that existing data simply support the relationship between repression-sensitization and self descrip-tion with no evidence of a relationship with deviant response tendency per se.

3. Anxiety

Since anxiety is given a prominent position in the dynamics of most psychological malfunctioning, a relationship between repression-sensitiza-tion and anxiety measures would be hypothesized. Extreme sensitizers might be expected to verbalize feelings of anxiety readily while control-ling the emotional experience itself while extreme repressers should avoid both the verbalization and the subjective experience. Presumably, only the verbal expressions of anxiety are amenable to investigation.

Ullmann and McReynolds (1963) reported a significant correlation of —.50 between scores on the F-I scale and ward ratings of anxiety on 53 neuropsychiatric patients. As predicted, anxiety is more characteristic of those on the facilitating or sensitizing end of the scale. As noted earlier, the R-S scale and the IPAT Self Analysis Form were found to correlate .76 in a group of 35 hospital patients (Lomont, 1964).

Instruments such as Taylor's Manifest Anxiety Scale and the Welsh Anxiety Scale are built from MMPI items just as the R-S and F-I scales are; further, there is even item overlap. Given these circumstances, a substantial relationship between repression-sensitization and anxiety could hardly be avoided. Ullmann and McReynolds (1963) reported correlations between the F-I scale and Welsh Anxiety of —.95 for psychiatric patients and —.85 for college students. Similarly, Joy (1963b) found a .91 correlation between the R-S scale and the MAS with under-graduates. Not only are sensitizers found to be more anxious, but scores on the repression-sensitization measure are about as highly correlated with these two anxiety scales as is possible with the instruments' reliabilities. A few of the implications of this somewhat disturbing finding are dis-cussed in the final section of this chapter.

4. *Emotional Stability and Social Approach-Withdrawal*

Weinberg (1963) carried out a factor analytic investigation of 37 variables in which the R-S scale was included along with word association scores, field dependency measures, and several personality scales. The subjects were 62 outpatients at the UCLA Medical Center Clinics. Ten factors were obtained; the first four (accounting for 70% of the variance) were labeled Emotional Stability *versus* Instability, Word Association Style, Field Dependence, and Social Approach-Withdrawal. The R-S scale loaded substantially on the first and fourth of these factors. Also loading on Factor I with the R-S scale were Impulsivity, Cyclothymia, Thinking Introversion, *Ma, Pd,* and two depression scales, while the opposite pole was represented by Achievement via Conformance and Responsibility. On Factor IV at the same pole as the R-S scale were Social Introversion, Depression, and sleeplessness while extraversion and rhathymia loaded in the opposite direction.

5. *MMPI Variables*

As with the anxiety measures, relationships between the R-S scale and other MMPI scales are complicated by overlapping items, identical test format, and so forth. In spite of these confounding factors, the findings reported by Joy (1963b) are of interest. With a group of 35 Texas undergraduates, he obtained significant correlations with a number of MMPI scales as shown in Table IV. Also shown are significant correlations found by Endler (1963) with a group of 42 undergraduate males

TABLE IV
CORRELATIONS BETWEEN R-S SCALE AND MMPI SCALES

	r	
	Joy (1963b)	Endler (1963)
MMPI Diagnostic Scales		
Depression	.72	.66
Hysteria	—.53	.01
Masculinity-Femininity	.50	—
Psychasthenia	.60	.92
Social Introversion	.84	—
Hypochondriasis	—.25	.52
MMPI Validity Scales		
Lie	—.61	—.38
F	.62	—
K	—.91	—.76
MMPI Special Scales		
Neuroticism	.75	—
Ego Strength	—.66	—

at York University. As a brief and tentative summary, the two investigations suggest that sensitizers, compared with repressers, tend to respond to the test as do members of the opposite sex, to give unusual or unconventional responses, to be depressed and discouraged, to be anxious and agitated, to be socially introverted, and neurotic. In addition, compared with repressers, sensitizers tend to be more cynical and self-debasing and to have less ego strength. Rather than a curvilinear relationship with adjustment, sensitizers tend to present a less emotionally healthy picture than repressers.

6. CPI Variables

The California Psychological Inventory (CPI) (Gough, 1957) was designed to measure 18 personality variables. In contrast to the psychodiagnostic orientation of the MMPI, the CPI centers on areas of behavior relevant to a normal population in four areas: Poise, Ascendancy, and Self-Assurance; Socialization, Maturity, and Responsibility; Achievement Potential and Intellectual Efficiency; and Intellectual and Interest Modes. On each variable, high scores indicate better adjustment than low scores.

TABLE V

REPRESSION-SENSITIZATION AND CPI VARIABLES

	r		
	Joy (1963b) 35 Undergraduates	Byrne, Golightly, and Sheffield (in press)	
		43 Males	48 Females
Dominance	—.52a	—.21	—.16
Capacity for Status	—.65a	—.29	—.22
Sociability	—.66a	—.32b	—.29b
Social Presence	—.47a	—.24	—.43a
Self-Acceptance	—.33	—.08	—.15
Sense of Well-Being	—.80a	—.43a	—.61a
Responsibility	—.37b	—.14	—.24
Socialization	—.38b	—.22	—.32b
Self-Control	—.60a	—.38b	—.53a
Tolerance	—.78a	—.45a	—.39a
Good Impression	—.67a	—.39a	—.52a
Communality	—.03	—.11	—.23
Achievement via Conformance	—.72a	—.46a	—.46a
Achievement via Independence	—.55a	—.33b	—.07
Intellectual Efficiency	—.55a	—.32b	—.52a
Psychological-Mindedness	—.41b	—.21	—.16
Flexibility	—.17	—.24	—.07
Femininity	.38b	.18	.10

a $p < .01$.
b $p < .05$.

Two investigations have included the correlation of each CPI scale with the R-S scale in groups of undergraduates. Joy (1963b) utilized 35 undergraduates of both sexes while Byrne, Golightly, and Sheffield (in press) obtained samples of 43 male and 48 female undergraduates. The coefficients are shown in Table V. As may be seen, all three samples yielded significant correlations between repression-sensitization and seven of the 18 CPI scales: Sociability, Sense of Well-Being, Self-Control, Tolerance, Good Impression, Achievement via Conformance, and Intellectual Efficiency. On the basis of the CPI manual, repressers and sensitizers would tend to be differentially described by others as in Table VI.

TABLE VI

Repressers	Sensitizers
1. Active, alert, ambitious, competitive, energetic, enterprising, forward, industrious, ingenious, productive, progressive, quick, resourceful, strict and thorough in their own work and in expectations for others, and valuing work and effort for its own sake	1. Apathetic, excitable, lacking in self-direction, leisurely, overly judgmental in attitude, passive, pessimistic about their occupational futures, quiet, submissive, unambitious, and unassuming
2. Broad and varied interests, capable, clear-thinking, high value on cognitive and intellectual matters, intellectually able, intelligent, original and fluent in thought, valuing intellectual activity and intellectual achievement, verbally fluent, versatile, and well-informed	2. Coarse, constricted in thought and action, conventional and stereotyped in thinking, narrow, opinionated, shallow, and shrewd
3. Calm, conscientious, deliberate, diligent, efficient, honest, inhibited, organized, patient, persistent, planful, practical, responsible, self-denying, sincere, slow, stable, thorough, and thoughtful	3. Awkward, cautious, confused, easily disorganized under stress or pressures to conform, easygoing, impulsive, insecure, lacking in self-discipline, stubborn, and uninhibited
4. Concerned with making a good impression, cooperative, helpful, informal, outgoing, sociable, and tolerant	4. Aggressive, aloof, apologetic, assertive, cool and distant in their relationships with others, defensive, detached, disbelieving and distrustful in personal and social outlook, irritable, overemphasizing personal pleasure and self-gain, overly influenced by others' reactions and opinions, retiring, self-centered, self-defensive, suggestible, suspicious, too little concerned with the needs and wants of others, and wary

Though these data are not entirely consistent with all of the findings with the R-S scale reported in this chapter, they present a relatively convincing picture of repressers as appearing better adjusted than sensitizers. This characterization is even more striking in Fig. 4, in which Byrne, Golightly, and Sheffield's (in press) subjects were divided roughly into thirds on the basis of their R-S scale scores (29 repressers, 29 sensitizers, and 33 neutrals) and their mean standard scores on the CPI plotted. The
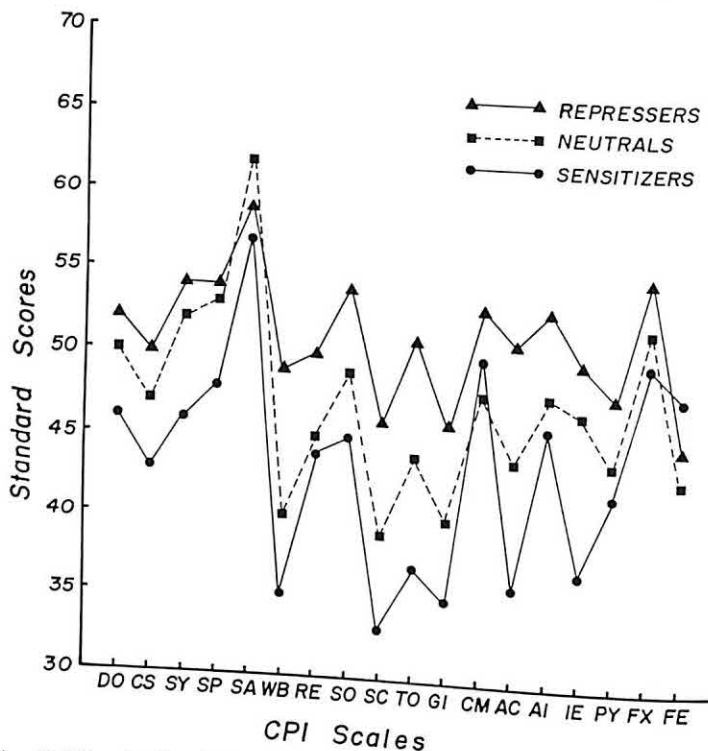
Fig. 4. California Psychological Inventory profiles of repressers, neutrals, and sensitizers. [From Byrne, Golightly, and Sheffield (in press).]

repressers may be seen to fall at or above the mean of Gough's standardization sample with the neutrals and sensitizers falling progressively lower on most of the scales.

7. Overcontrol

Megargee (1964) has presented evidence supporting the hypothesis that extremely assaultive criminals (e.g., murderers, those who commit assault with a deadly weapon) appear less aggressive in their everyday behavior and in test responses than do other criminals or normals. The picture is one of a chronically overcontrolled individual who suddenly

breaks forth with extreme violence, often to the surprise of those who know him.

In subsequent work, Megargee and Mendelsohn (1964) developed and cross-validated an MMPI scale which discriminates extremely assaultive offenders from other criminals and from normals. Since repression and denial are among the mechanisms utilized by the overcontrolled individual, they hypothesized a negative relationship between this new scale and the R-S scale. With a sample of 14 extremely assaultive, 28 moderately assaultive, and 44 nonviolent criminals plus 50 male undergraduates, the two scales were found to correlate $-.41$ ($p < .01$) as predicted. Thus, extreme repressers are seen to fit into the pattern of chronic overcontrol which includes rigidity, conventionality, moralistic views, and (at least in the criminal group) a greater probability of psychosis (Megargee and Mendelsohn, 1964).

8. Conclusions

The proposition that the two extremes of the repression-sensitization dimension represent different but equally maladjusted ways of responding to anxiety and conflict is not consistent with the majority of evidence now available. The only supportive data are the findings which indicate greater variance on the defensive measure in a neuropsychiatric than in a college population and the correlation with the Megargee and Mendelsohn measure of overcontrol. All other findings strongly suggest a linear relationship between sensitizing defenses and maladjustment. Scores on the R-S scale are positively related to tendencies toward alcoholism, the number of deviant responses given on an adjective check list, anxiety, emotional instability, social withdrawal, and a number of MMPI and CPI scales which indicate various types of psychological maladjustment. Those on the repression end of the continuum were found to be higher on the hysteria dimension in one investigation but not in another. Repressers also receive higher scores on a measure of ego strength. The weakness of the majority of this evidence is that paper-and-pencil tests are utilized as indicators of maladjustment, and repressers by definition are those who respond to such instruments by denying that anything is wrong with them. Whether they are presenting an accurate picture of their emotional health will ultimately have to be determined by means other than asking them again on a different paper-and-pencil test.

D. Additional Antecedents, Correlates, and Consequents

1. Tendency to Give Socially Desirable Responses

In 1953, Allen Edwards published an article which was to prove the precursor of a body of fruitful and at times disturbing research dealing

with the effect of the social desirability of the content of test items. When he obtained social desirability scale values for a series of questionnaire items, the correlation between this value and the proportion of subjects agreeing with the item content was .87. This substantial relationship has been confirmed by subsequent investigations.

A related, but somewhat different concept, is Edwards' (1957) social desirability hypothesis, which simply proposes the existence of individual differences in the tendency to give socially desirable responses to questionnaire items. He has built a 39-item Social Desirability (SD) scale to measure this tendency. In a number of investigations, scores on the SD scale have been found to correlate highly with scores on many personality measures, including most of the MMPI scales (Edwards, 1959).

It would seem reasonable to expect that the R-S scale would also show a relationship because it is made up of MMPI items. In addition, individuals who utilize repressive mechanisms should stress socially desirable characteristics while sensitizing individuals should emphasize socially undesirable characteristics. With a group of 35 undergraduates, Joy (1963b) found a correlation of —.91 between the Edwards SD scale and the R-S scale.

Liberty, Lunneborg, and Atkinson (in press) conducted a factor analytic investigation in which several personality measures were given to 150 male students. The MMPI was scored for 62 scales including the R-S and F-I scales. Factor analysis of the intercorrelation matrix yielded 10 factors. Most of the variance was accounted for by the first three factors: social desirability, acquiescent set, and lie. Both repression-sensitization measures primarily loaded on the first of these factors. Factor I was labeled social desirability because loadings of scales on that factor correlated .93 with proportion of keyed SD items in the scales and because the Edwards SD scale was a relatively pure measure of that factor.

In a similar investigation, Wiggins (in press) factor-analyzed 30 scales from eight tests, including the F-I scale, with a group of 137 University of Illinois students. Of the six factors identified, the F-I scale had its highest loading on the first or social desirability factor.

Another approach to social desirability has been that of Crowne and Marlowe (1960), who point out that the Edwards SD scale is characterized by items containing primarily pathological implications. In a normal population, a high score, then, may mean that the subject is attempting to give the most socially desirable responses or that he is actually free of pathological symptoms. To remedy this problem, they built a scale consisting of items dealing with culturally sanctioned behaviors that are not likely to occur. The resulting Crowne–Marlowe SD scale was found to correlate .35 ($p < .01$) with the Edwards SD scale.

In both of the factor analytic studies indicated above (Liberty et al.,

in press; Wiggins, in press), the Crowne–Marlowe SD scale loaded high in a different factor from the Edwards SD scale and from the repression-sensitization measures. In unpublished research at the University of Texas, however, a significant negative correlation ($r = -.37$, $p < .01$) between the Crowne–Marlowe SD scale and the R-S scale was found in a group of 115 students.

2. Authoritarianism and Dogmatism

In characterizing the personality dynamics of authoritarians, Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) stressed the mechanism of repression as one of the basic components. A correlation of $-.40$ ($p < .01$) between the R-S and F scales has been reported (Byrne, 1961). The authoritarian measure was a 10-item scale composed of five of the original F-scale items and the five reversed items of Couch and Keniston (1960). Subsequent research at Texas (Byrne and Bounds, 1964) has found that there is no relationship between response to the Couch and Keniston items and scores on the original F scale. Therefore, the meaning of the reported correlation is ambiguous at best.

An unpublished study by Friedman (1963) used a 32-item F scale (27 original items and 5 successfully reversed ones). This scale correlated .28 with the R-S scale; with 33 subjects this coefficient was not significant. In another unpublished study, Byrne and Goldberg (1962) used the original F scale and found a nonsignificant correlation of .13 between authoritarianism and repression-sensitization in a group of 76 students.

Rokeach (1960) in his work on open and closed mindedness has developed a Dogmatism scale designed to tap rigidities of both the political left and right. In addition, his Opinionation scales directly tap extreme political attitudes. In the Byrne and Goldberg investigation, in addition to the F scale and the R-S scale, Rokeach's Dogmatism, Left Opinionation, and Right Opinionation scales were administered to the 76 subjects. Significant positive correlations were found between the R-S scale and dogmatism ($r = .36$, $p < .01$) and left opinionation ($r = .33$, $p < .01$), while repression-sensitization was not related to right opiniona-tion ($r = .05$). Joy (1963b) reported a correlation of .55 ($p < .01$) between R-S scale scores and scores on the MMPI Prejudice Scale.

In unpublished research at York University, Sermat (1962) sought a relationship between repression-sensitization and belief in the likelihood of atomic war. Scores on this item correlated .38 ($p < .05$) with the R-S scale for 30 male students but only .02 for 30 female students.

3. Other Personality Variables

In the Weinberg (1963) factor analytic study cited earlier, 62 out-patients were given a series of measures including the R-S scale, the

DONN BYRNE

Myers–Briggs Type Indicator, and Guilford's Inventory of Factors STDCR. As shown in Table VII, seven of the nine variables measured by these tests were significantly related to repression-sensitization. On the basis of these relationships, sensitizers would tentatively be described as more introverted, neurotic, emotionally maladjusted, and as responding

TABLE VII

CORRELATIONS OF THE R-S SCALE WITH THE MYERS-BRIGGS AND GUILFORD TESTS[a]

	r
Myers-Briggs Type Indicator	
Extraversion	$-.42^b$
Sensation	$-.21$
Thinking	$.16$
Judgment	$.28^c$
Guilford Inventory of Factors STDCR	
Social Introversion; Seclusiveness	$.58^b$
Thinking Introversion; Reflectiveness	$.44^b$
Depression; Unhappiness; Pessimism	$.81^b$
Cycloid Disposition; Emotional Instability	$.77^b$
Rhathymia; Carefreeness; Happy-Go-Lucky Disposition	$-.28^c$

[a] After Weinberg (1963).
[b] $p < .01$.
[c] $p < .05$.

on the basis of perception rather than judgment. Repressers, on the other hand, appear to be extraverted and well adjusted, to have self-control, and to utilize judgment in preference to perception.

4. Intellectual Ability

In delineating the meaning of test variables, the absence of certain relationships may be as meaningful as the presence of other relationships. Four investigations have included the repression-sensitization variable and an intellectual measure. Nonsignificant correlations have been found between the R-S scale and the Shipley–Hartford Scale with 132 undergraduates ($r = -.15$) (Byrne, 1961) and with 35 hospital patients ($r = .07$) (Lomont, 1964), standard scores on a college entrance test with 26 male undergraduates ($r = .25$) (Byrne, 1961), and college ability as measured by the SCAT with 30 male students ($r = .26$) and 30 female students ($r = .03$) (Sermat, 1962), and with 42 male students ($r = .24$) (Endler, 1963). In addition, Endler (1963) failed to find a relationship between R-S scale scores and High School Grade Point Average ($r = .09$) or College Final Grade Average ($r = -.02$) with 42 first-year men at York University.

The only suggestion of a relationship here is that of a small positive correlation between repression-sensitization and intelligence for males.

Following up this possibility, the author divided the 132 students in the Shipley–Hartford study into subgroups of 60 males and 72 females and computed the correlations separately. The r's of −.18 and −.11 were not significantly different from zero. On the basis of existing data, then, the R-S scale and measures of intellectual ability appear to be independent.

5. Parental Antecedents

In an experiment reported in 1957, Dulany successfully "taught" patterns of repression and sensitization to subjects. With geometrical figures as the stimuli, perceptual repression was induced by punishing subjects' responses to a threatening stimulus, thus allowing competing responses to be reinforced via anxiety reduction. Perceptual sensitization was induced by punishing competing responses while allowing perceptual responses to the threatening stimulus to occur and hence receive reinforcement.

Generalizing from these findings, Byrne (in press) hypothesized that parental influences on the development of repressive and sensitizing defenses should be similar to Dulany's laboratory procedures. That is, in childhood repressers experienced punishment in their attempts to express conflictful impulses while sensitizers were permitted to express such material, at least at a verbal level. Thus, child-rearing attitudes falling along a permissive-restrictive dimension should be related to the development of sensitizing *versus* repressive mechanisms in the offspring. In three investigations, R-S scale scores were correlated with responses of mothers on the Parent Attitude Research Instrument (Schaefer and Bell, 1958), subjects' responses on the Hereford Parent Attitude Survey (Hereford, 1963), and subjects' perceptions of their mother's responses to the Stanford Parent Attitude Questionnaire (Winder and Rau, 1962). The findings were that a relatively small but consistent group of the parental variables were related to repression-sensitization, as summarized in Table VIII. Though the findings must be cross-validated, the repressers indicate a home atmosphere characterized by permissiveness, acceptance, and confidence. Their mothers were consistent and high in self-esteem while the two parents had a positive affective relationship with one another. The sensitizers, on the other hand, report a restrictive and rejecting home, and a lack of confidence in assuming the role of parent. Their mothers were inconsistent, low in self-esteem, and had a negative affective relationship with the father. It is obvious that the findings are quite different from those which were hypothesized.

Another investigation of represser-sensitizer differences in their descriptions of their parents was carried out by McDonald (in press). The subjects were 177 unmarried pregnant women, ranging in age from

14 to 39. Most of them were from lower middle class homes. On the R-S scale, the group's mean and standard deviation were comparable to those of female college students (Byrne, 1961). McDonald also identified subjects as expressors on the basis of MMPI scores: 2 Pd–(Welsh repression + Hy denial). Subjects were classified into four extreme groups consist-

TABLE VIII
REPRESSION-SENSITIZATION AND CHILD-REARING ATTITUDES[a]

Repressers	Sensitizers
Males	
1. Mother permissive about sexual behavior	1. Mother suppressive about sexual behavior
2. Mother perceived as having high self-esteem	2. Mother perceived as having low self-esteem
3. Mother perceived as having high expectations of son achieving masculine role	3. Mother perceived as *not* having high expectations of son achieving masculine role
Females	
1. Mother strict	1. Mother permissive
2. Mother does not expect deification of parents	2. Mother expects deification of parents
3. Feels confident in parental role	3. Does not feel confident in parental role
4. Mother perceived as accepting	4. Mother peceived as rejecting
5. Mother perceived as nonpunitive and not using physical punishment	5. Mother perceived as punitive and as using physical punishment
6. Mother perceived as having high self-esteem	6. Mother perceived as having low self-esteem
7. Father-Mother relationship perceived as positive	7. Father-Mother relationship perceived as negative
8. Mother perceived as consistent	8. Mother perceived as inconsistent

[a] After Byrne (in press).

ing of 34 repressers, 14 sensitizers, 23 expressors, and 34 expressor-sensitizers. Each subject filled out the Interpersonal Check List for self, mother, and father. The major emphasis of the investigation was on the hostility items. As would be expected, repressers indicated significantly fewer of the hostility items in the self-ratings than each of the other three groups. With respect to ratings of parents, repressers attributed significantly less hostility to their fathers than did any of the other groups. On the mother ratings, repressers and sensitizers were not significantly different, but repressers did attribute significantly less hostility to their mothers than either the expressors or the expressor-sensitizers. When the

data were analyzed on the basis of self-parent discrepancies, the repressers consistently indicated the least discrepancy (especially with their mothers) while the expressor-sensitizers indicated the greatest discrepancy.

6. Interpersonal Behavior

Any pervasive personality variable, such as repression-sensitization, is potentially an important determiner of some aspects of interpersonal behavior. An individual's socially relevant motives, his perception of others, his response to the demands of group situations, and his effect on others are likely to be in part a function of his characteristic defense modes.

Altrocchi (1961) asked each member of a group of student nurses (46 repressers and 46 sensitizers selected from a much larger group) to describe herself and three randomly selected classmates in terms of the items on Leary's Interpersonal Check List. He found that the assumed dissimilarity between self and others was significantly higher for sensitizers than for repressers, regardless of the degree of actual similarity-dissimilarity. Further analysis of the data indicated that the findings could be explained on the basis of the familiar represser-sensitizer differences in self concept rather than differences in the perception of others.

Altrocchi's Expressor Index [2 Pd–(Welsh $R + Hy$ denial)] has proven to be a useful means of differentiating two types of sensitizing individuals. Altrocchi and Perlitsh (1963) hypothesized differences in attribution of hostility to self and others by sensitizers, repressers, expressors, and expressor-sensitizers. The 101 subjects were selected from a group of 296 senior nursing students at Duke University. Attribution of hostility was measured by the Love scores of Leary's Interpersonal Check List. It was found that expressor-sensitizers attributed more hostility to themselves than did repressers or expressors, expressors attributed more hostility to others than did repressers, and more hostility was attributed by others to expressors than to sensitizers. Altrocchi and Perlitsh (1963, p. 817) suggest,

"As in previous studies, the present results demonstrate no difference between repressors and sensitizers in attribution of hostility to others, but add the expected finding that neither of these two groups is seen by others as particularly hostile. Repressors do not seem to think that feelings of anger or hostile behavior are salient aspects of their functioning . . . , and their associates in this study seem to agree; clinicians, however, do not. It remains to be demonstrated experimentally whether clinicians are correct in believing that hostile impulses are aroused as readily in repressors as in other people, but are simply repressed."

A subsequent investigation (Altrocchi et al., 1964) suggests that there

are sex differences in the attribution of hostility to others in addition to differences based on defenses.

In correlating the R-S scale with various MMPI scales, Joy (1963b) found repression-sensitization to correlate −.87 with leadership. Partly on the basis of this finding, he instituted an investigation (Joy, 1963a) with 30 three-man groups, each composed of one represser, one sensitizer, and one neutral (a subject scoring in the middle quarter of the scale). The groups discussed a standard human relations problem, and then each subject rated himself and the other two group members on a series of variables. Half of the groups were given a task orientation and half an ego threat orientation. The only significant finding based on ratings by others was that sensitizers were chosen less often as desirable work partners than either repressers or neutrals. He also found that repressers rated themselves as showing more concern for maintaining friendly relations than did the sensitizers; neutrals rated themselves as showing more leadership ability than did the two extreme groups.

Turk (1963) investigated several aspects of dyadic relationships in teams composed of a student nurse and a student physician. The nurses were classified as sensitizers or repressers, but Turk preferred to label the former as having low presentational conformity and the latter high presentational conformity. The physicians were not measured on this variable but were assumed to vary randomly in assignment to the two types of nurses. Each team provided three weeks of professional health care in the outpatient clinic of a university hospital. After the work period, each subject completed a questionnaire, including questions concerning their enjoyment in working on the task and the amount of enjoyment attributed by teammates to each other. It was found that the correlation between the enjoyment which the physician attributed to the nurse and that which she herself expressed was significant only in the groups with sensitizing nurses: nurse's self-rating and physician's rating of nurse correlated .69 ($p < .001$) in the sensitizer groups and −.17 in the represser groups. While the physicians were accurate only in the sensitizer groups, assumed similarity in enjoyment was significant only in the groups with repressing nurses. Physician's self-rating and his rating of the nurse correlated .80 ($p < .001$) while nurse's self-rating and her rating of the physician correlated .62 ($p < .01$) in the repressing groups. In the sensitizing groups, these two correlations were .35 and .30. One interpretation of these findings is that accuracy of interpersonal perception is most likely to occur when the perceptual target is a sensitizer who exhibits consistency in expressions of affect in overt behavior and in self-ratings. With repressers, the overt communication of feeling and the self-ratings

were apparently inconsistent, thus making possible distortion via assumed similarity.

Silber and Baxter (1963) have supplied preliminary data from an investigation in which repressers and sensitizers were compared on several aspects of a verbal conditioning task. It seems possible to interpret their findings as reflecting differential responses to an interpersonal situation. They attempted to condition 52 undergraduates at the University of Kentucky to respond with either thought or motor-action verbs. On the basis of R-S scale scores, each subject was classified as a sensitizer or a represser. On index cards, the pronoun "he" appeared at the top with three alternative words (thinking, motor-action, and verbal activity) at the bottom. The subjects were instructed to make a two-word sentence, using the pronoun and one of the verbs. There was an operant level phase of 20 trials, an acquisition phase of 100 trials ("Mmm-hmm" for a thought verb for half the group and for a motor-action verb for the other half), and an extinction phase of 40 trials. A postconditioning interview was conducted to determine the subjects' awareness of the experimental conditions and their estimate of the frequncy of verbal reinforcements by the experimenter. Among the results relevant to repression-sensitization, Silber and Baxter found no differences between the groups in operant levels for the verbs, a significant interaction between repression-sensitization and awareness (aware repressers gave the greatest number of responses and aware sensitizers the smallest number) during acquisition, and greater extinction effects for repressers than sensitizers. Finally, compared with sensitizers, repressers overestimated the number of reinforcements emitted by the experimenter. It might be concluded that in this type of social situation repressers differ from sensitizers in that they respond more to the implicit demands of the other person when they realize what those demands are. They more readily respond with a given class of words when the experimenter offers approval for this and more readily cease this type of response when the experimenter withdraws his approval. In addition, they misperceive the situation such that they remember receiving a greater number of social reinforcements than was actually the case. Sensitizers, in contrast, appear to reject or resist the demands of the situation. If this interpretation of the findings is an accurate one, repressers and sensitizers should not differ in response to an operant conditioning task which does not involve an interpersonal context.

Shrader (1963) at the University of Massachusetts selected a group of 72 male subjects from a pool of 194 who had taken a self-ideal discrepancy questionnaire and the R-S scale. The experimental task was a modified Ames Thereness-Thatness apparatus in which subjects made

size and distance judgments of three full-face, life-sized photographs of college males. Each was described as a superior achiever in intellectual, social, or physical activities. Shrader's findings with respect to the defense groups were that repressers perceived the stimuli as larger and closer than the sensitizers and that they were more variable in their responses. The difference between repressers and sensitizers decreased as the stimulus distance increased. Shrader interpreted these results as meaning that the repressing individuals were more suggestible in responding to the experimenter's instructions and that they were more labile in responding. The differences decreased when the stimulus was farther away because of the reduced threat qualities at the greater distance.

7. Obstetric Complications

The effects of emotional stress and of personality differences in dealing with stress have been studied in connection with various obstetrical complications. Several research projects are under way to determine the part played by psychological variables in abortion, fetal mortality, labor time, and mental retardation. Recent findings suggest that differences along the repression-sensitization dimension are related to the probability of occurrence of some of these abnormalities.

McDonald, Gynther, and Christakos (1963) obtained data on 86 patients from the Obstetrical Clinic at the Medical College of South Carolina. Each patient was given the IPAT Anxiety Scale, the Kent EGY, and the MMPI, which was scored for repression-sensitization. On the basis of pregnancy, delivery-room, and postpartum records, the patients were classified as normal ($N = 42$) or abnormal ($N = 44$). The latter consisted of disorders of the gestation period, developmental abnormalities observed at birth, and delivery irregularities. The two groups of patients did not differ significantly in infant birth weight, labor time, Kent EGY scores, or age. They were, however, different at beyond the .01 level of significance in anxiety and repression-sensitization. The abnormal birth group was more anxious and more sensitizing. In the normal group, R-S scale scores were comparable to those of university coeds (Byrne, 1961) while patients in the abnormal group had a mean score higher than 85% of the women in the university sample.

One of the questions which arises from that finding is whether the high sensitization scores represent the characteristic functioning of these women in the abnormal group or whether they represent a special condition induced in part by the stress of pregnancy. McDonald and Parham (in press) investigated 160 unmarried women before and after the birth of their first child. The MMPI and Kent EGY were administered at the beginning of the seventh month of gestation; shortly after delivery (7–10 days) the

MMPI was readministered. As in the previous investigation, patients were divided into groups of 79 with normal and 81 with abnormal births. Comparison of test scores before and after delivery revealed that patients had significantly lower R-S scale scores (about 8 points) after the the baby was delivered. Comparison of the normal and abnormal groups indicated that they did not differ in IQ or age. Both groups were significantly less sensitizing after delivery, but at both periods the abnormal group was significantly more sensitizing than the normal group. They differed by 5.50 points before delivery and by 1.65 points after delivery. As the authors indicate, the reasons for the differences in ego defenses in the normal and abnormal groups and the shift toward more repressiveness after delivery are not clear.

8. Conclusions

To date, a number of relationships have been established between repression-sensitization and other personality variables. That the Social Desirability variable is negatively related to the repression-sensitization dimension is well established, but the meaning of this relationship is open to more than one interpretation. The R-S scale appears to be unrelated to authoritarianism, right opinionation, or intelligence, but positively related to dogmatism, left opinionation, and prejudice. A theoretical explanation for the latter three unexpected findings has not been offered. There are, on the contrary, quite reasonable grounds for expecting that individuals who utilize repressive defenses would be prejudiced, dogmatic, and opinionated. To find that it is those with sensitizing defenses who tend to be high on those three dimensions will require reevaluation of existing formulations. The substantial relationships between repression-sensitization and various measures of introversion-extraversion raise the possibility of interconnections between the type of work described in this chapter and that of Eysenck.

The attempts to extend knowledge about the repression-sensitization dimension by investigating its antecedents in child-rearing attitudes and its consequents in interpersonal behavior and somatic functioning represent a potentially important type of research with personality variables.

The area of personality shows increasing maturity as the antecedents and consequents of individual differences become known. With respect to repression-sensitization, it may be fair to state that *nothing* is yet known about its antecedents. The reported investigations can be interpreted most parsimoniously as relationships between the R-S scale and particular types of self-report measures. The fact that repressers and sensitizers report different things about their parents is informative, but it quite probably provides no evidence about actual parental differences in

attitudes or behavior. Some of the findings concerning consequents in interpersonal behavior are subject to the same comments. However, the results indicating that sensitzers were not chosen as work partners, that perception by outsiders of the feelings of sensitizers was more accurate and less distorted than the perception of the feelings of repressers, that repressers and sensitizers differ in their responses in an operant conditioning task, and that they differ in judgments of size and distance of photographs, give promise of fruitful research possibilities in this area. Similarly, when sensitizing women are found to have greater difficulty in giving birth to their infants, it is clear that more is involved in represser-sensitizer differences than simply verbal response sets on paper-and-pencil tests. A variable is "good" to the extent that it is related to other variables; repression-sensitization appears to qualify for that designation.

III. Repression-Sensitization: What Are the Next Research Steps?

The research findings with the repression-sensitization dimension lead to two major questions about subsequent developments in this area. First, is the present conceptualization of this variable as involving individual differences in approach and avoidance defenses a reasonable one or does some other concept better fit the data? Second, what are the next steps in theory and research to be taken with this variable?

A. What Is Being Measured by the R-S Scale?

It seems prudent to view any investigation which uses a personality test as constituting, at least in part, a construct validity study. In the traditional psychometric sense, there is literally no way of validating or invalidating a personality test. Instead, the instrument may be evaluated with respect to its measurement consistency or reliability, its relationships with other personality variables may be determined, and its antecedents and consequents may be specified. It may then be discarded as an unreliable measure or shelved by researchers because it lacks relational fertility. The nearest thing to invalidation that can occur is a change in the name of the test. As the author (Byrne, 1964, p. 51) has suggested elsewhere:

"Another way of asking the validational question is, 'What is the best name to apply to the responses measured by this instrument?' As simple (or simple-minded) as this question appears, much of the controversy centering around many tests resolves itself to the naming problem. Hopefully, the name designated for any measured behavior may be thought of as tentative. Perhaps the following qualifying statement should be assumed wherever the name of a test appears: 'in view of the item

content of this test and its relationships with other variables, at this time the instrument appears to be conceptualized as a measure of ——'."

If one is able to tolerate a certain amount of confusion and ambiguity, the characteristics of the variable being measured may be made more clear after a sufficient period of research activity. In the research reviewed in the previous section, for example, the dimension which we have discussed as repression-sensitization has been designated by others as facilitation-inhibition, affector-effector, and presentational conformity. In addition, evidence has been presented which suggests that the labels self concept, anxiety, neuroticism, or social desirability would be reasonable alternates. To assert that the R-S scale is *only* measuring response set or *really* measuring anxiety, as if these new labels somehow brought about conceptual clarity or as if existing data were no longer meaningful, is a sort of word magic.

Do the labels or the test names matter then? Would it not be wise to use nonsense syllables or zip codes as the names of personality variables and thus side-step the problem? This approach seems unwise because the name of the variable and the surplus meaning attached to that name are likely to have a great influence on the type of research that is conducted and on the theoretical structure into which the findings are placed. Differences in labels and theory mean that data may be sought which would otherwise seem irrelevant. Farber (1964, p. 36), for example, points out that the fruitfulness of the MAS is its supposed relevance as a measure of drive in Hull–Spence theory. Even though the MAS and Edwards Social Desirability Scale correlate —.84, he suggests that it is doubtful whether investigators using the latter variable would have conducted studies "of eyelid conditioning, or psychophysical functions, or the steepness of stimulus generalization gradients, or paired-associate learning."

As with the R-S scale, users of the MAS point to the very high relationships reported among the MAS, the *Pt* and *K* scales of the MMPI, and the Social Desirability Scale of Edwards. Farber (1964, p. 30) asks,

"What, then, does the test measure?

"In one sense, this is a trivial question, and requires but a trivial answer, though, unfortunately, one that constructors and users of tests sometimes fail to see, namely, that giving tests different names does not guarantee that they reflect different characteristics, and giving them the same names does not necessarily mean they reflect the same determinants. If all these highly interrelated measures are related to all other kinds of behaviors in the same way, they measure the same thing, regardless of their labels.

"Occasionally, however, this confusion among the characteristics inferred from behavior is not merely nominal. For instance, height and

weight are highly correlated in the general population, yet no one supposes they are merely different names for the same thing. What if defensiveness and desire to make a favorable impression are independent, but nevertheless empirically related in a given population? How, then, could we decide whether they reflect different organismic states or processes? Or better, if the one measure is related to some other mode of behavior, which hypothetical variable is responsible for the relation?

"These are not trivial questions, and their answers are not easily come by. One kind of answer is simply the observation that there is never any guarantee in science against the inaccurate identification of determinants."

Farber (1964, p. 31) goes on to suggest:

"It is usually possible, given skill and patience, to tease out ever finer specifications of the variables entering into behavioral laws. If we consider response-defined variables, by careful selection of cases, statistical correction, or the elimination of errors of measurement, we can frequently decide which of several variables, even though they be highly correlated, determine the form of the relation under consideration. If it turns out, as it frequently does, that a given behavior is a joint function of several response variables, the circumstance is one quite familiar to experimentalists. For instance, no one is unduly disturbed by the finding that the probability of drinking is a function of both dryness of the mucous membranes and general water deficit. The additional fact that these two variables are themselves frequently related, under many conditions, is a complication, to be sure, but not an insuperable one.

"All we can reasonably ask in regard to either experimental or differential variables is that they be specified as precisely as possible. We should then be prepared to discover, soon enough, that more precise and detailed specifications are necessary."

With the repression-sensitization measures, the evidence would appear to be rather clear-cut that the defense scales and the Edwards SD scale are essentially interchangeable, two different names for the same variable. Even though this interpretation seems obvious, Ullmann and McReynolds (1963) have presented data which indicate otherwise. As reported earlier, they found a —.50 correlation between the F-I scale and ward ratings of anxiety in a group of neuropsychiatric patients. This relationship was also computed with scores on the Edwards SD scale held constant by means of partial correlation. The resulting coefficient was still a significant one ($r = -.38$, $p < .01$), which suggests clearly that something is being measured by the defense test in addition to social desirability.

Given such a finding plus the many studies which contribute to the construct validity of the instrument, at the present time the label

"repression-sensitization" does not appear inconsistent with what is known about the R-S scale. There is as yet no compelling reason to re-name the test.

B. CONSIDERATIONS IN BUILDING A THEORY OF REPRESSION-SENSITIZATION

The history of the field of personality may be seen as containing the discarded remains of a vast number of variables investigated over the years—each measured more or less adequately, each the object of a flurry of research interest, and each passing out of favor. Except at the methodo-logical level, it is difficult for even the most hardened optimist to view all of this activity as representing the forward motion of behavioral science with each generation more enlightened than the previous one. If this characterization can be accepted, it may be helpful to consider the nature of research and theory building at a more general level than is represented by the specific problem of repression-sensitization. What is involved in pursuing an area of research such as that described in the present chapter and what would it mean to proceed with the building of a theory of repression-sensitization?

1. General Considerations

The goal of that branch of science which is psychology is the build-ing of theories which permit the accurate prediction of behavior. Beyond a general acceptance of "the scientific method," however, there are a diversity of positions taken by psychologists with respect to the strategy and tactics best suited to attaining that goal. One of the differences among psychologists is in their conception of the role of theory in research. This question includes the assumptions which underlie a theory and guide its general framework (e.g., self theory with a phenomenological base *versus* S-R theory with a mechanistic determinism), the point at which theory assumes importance in research (e.g., prior to the design of the first experiment *versus* only after a long period of data gathering), the type of theory which is constructed (e.g., artistic description *versus* mathe-matical relationships among operationalized variables), and the breadth of the theory (e.g., a global framework which encompasses all of behavior *versus* relatively narrow theories tied to a specific series of behaviors). Each of these issues will be considered briefly.

2. Pretheoretical Assumptions

Behavioral science seems to mean quite different things to different individuals. One of the best descriptions of the kinds of assumptions held by psychologists in the experimental, behavioristic, operational tradition

has been presented by Farber (1964). The basic assumption is that be-
havior is a function of its antecedents (i.e., determinism). Farber (1964,
p. 32) says,

"What they come to, in a word, is an emphasis on the formulation
of empirical laws and the analysis of the variables comprising them. As
we have noted, the assumption that it is the business of behavioral science
to explicate the relations between objectively defined environmental and
behavioral events is in itself no theory, but rather a metatheoretical or
pretheoretical preference. Those who adopt this approach, especially if
they have an interest in the phenomena of learning, are likely to use the
terms 'S' and 'R' to refer, respectively, to the environmental and be-
havioral events, and the familiar formula 'S-R' to indicate a relation
between these two classes of variables."

From this frame of reference, then, the repression-sensitization vari-
able should be placed within a framework of empirical laws. It will be
possible to predict individual differences along this dimension when its
antecedents are known and to predict the effects of these differences on
other behavior when its correlates and consequents are known. Our goal
is not the *Verstehen* of the Romantic movement which would purport to
provide an understanding of repression-sensitization, but rather the goal
is to place this variable in a predictive framework as Holt (1962) has
cogently argued. The end point will be the specification of lawful rela-
tionships among variables rather than a colorful literary description of
repressers and sensitizers.

3. *Time for Theory Building*

In a sense, it is virtually impossible to design any piece of research
and execute it without the operation of at least an implicit theory. Vari-
ables are seldom selected through the use of tables of random numbers.
Beyond one's pretheoretical assumptions, there are expectations or
hypotheses or hunches or general notions about relationships between
X and Y. With the R-S scale, the author's initial expectations may be
summarized briefly:

(1) Individuals may be placed along a continuum with respect to
their characteristic learned response to threatening stimuli; avoiding
mechanisms define one end of this continuum and approaching mecha-
nisms the other.

(2) A curvilinear relationship should be found between an indi-
vidual's position on this continuum and various indices of his maladjust-
ment.

All of the research on repression-sensitization which has been con-
ducted by the author and much of the research conducted by others

follow from these two propositions. There is a great deal of support for the first and only modest support for the second. In addition to these findings, there are a number of other relationships which have been investigated and reported. Is it now time to tie all of these findings together into a more inclusive theory of repression-sensitization? It would certainly be the "socially acceptable" thing to do in terms of the mores of the field of personality. Perhaps, however, it is time to offer a dissenting opinion. It is instructive to compare work in personality with that of other scientists, including some of our psychological colleagues. Among the basic differences between the work of astronomers, biologists, physicists, chemists, and even learning theorists and the work of personality psychologists is a superficially simple one. In the other fields, the number of different variables is relatively small and there is relatively general agreement within the field as to what those variables are. Individual scientists may differ in what they do with the variables and certainly in how they interpret the findings, but there is a great deal of agreement otherwise. In personality research, the mode is an unenviable creativity with respect to both stimulus and response variables. It is difficult to find as many as half a dozen studies in our journals which utilize the same independent or dependent variables. It would be a disheartening exercise to survey and list the number and variety of different defining operations for hostility, anxiety, threat, or any other variables widely studied in personality.

At some point in the development of our field, this way of proceeding must change. Personality psychologists must find their equivalents of planetary motion, speed of a falling body, running speed in a straight alley runway, and frequency of the lever-pressing response. The next step will be the establishment of empirical laws which indicate the relationships between and among these variables. A law relating repression-sensitization to amount of anxiety aroused in response to threat, for example, will not arise from a series of investigations each of which involves the use of different operations to define anxiety and threat, each significant at least at the .05 level, and each conducted only a single time. When a series of such laws are in fact established, it will be the appropriate time to devise theories in an attempt to organize these laws into more inclusive units.

In the meantime, it is possible to formulate simple predictions and hypotheses based on generalizations from observation, deductions from other theories, and inductions from previous research. Any attempt to get more elaborate in our preliminary theorizing would seem to be an unfruitful endeavor on the basis of past experience. Many possibilities for research have been raised by those working with repression-sensitization and interpersonal behavior, obstetric complications, and maladjust-

ment, for example. There would seem to be little advantage in jumping from these interesting possibilities to a formal theory without pausing first to obtain much more extensive data.

4. Type of Theory

For many psychologists, the field of personality has been synonymous with broad descriptive formulations employing anecdote, allusion, meta- phor, and analogy while eschewing operational definitions and mathe- matical formulas. Obviously, the typical investigator in personality re- search does not match that characterization. It seems equally clear that the individual research efforts, the emerging laws, and the future theories of personality will bear considerably more resemblance to physics than to psychoanalysis.

For most behavioral scientists, these points are either too absurd to consider seriously or too obvious to bear further elaboration.

5. Breadth of Theories

Again, existing personality "theories" are characterized by their breadth. Typically, an attempt is made to explain all of man's behavior in terms of a given series of descriptive formulations. Moreover, the crowning achievements of other fields tend to be theories which are ex- tremely broad and hence able to unify a multitude of established findings: Geocentrism, evolution, and $e = mc^2$ are familiar examples.

Perhaps it would be prudent for personality psychologists to attempt somewhat less ambitious theory building at this point in the development of knowledge about human behavior. The necessary combination of indi- vidual genius and a suitable body of data apparently has a fairly low base rate. Instances of the former without the latter have been charac- teristic of personality psychology, and Sigmund Freud is the outstanding example. Instances of the latter without the former are met daily in most sciences. One of the reasons for the rarity of the successful broadly en- compassing theory is that error is increasingly easy as one moves from the data on which the laws were built. As Rogers (1959, p. 191) puts it,

"A slight error in a theory may make little difference in providing an explanation of the observed facts out of which the theory grew. But when the theory is projected to explain more remote phenomena, the error may be magnified, and the inference from the theory may be com- pletely false . . . . Thus, every theory deserves the greatest respect in the area from which it was drawn from the facts and a decreasing degree of respect as it makes predictions in areas more and more remote from its origin."

What, then, might our theoretical goals be? Elsewhere the author suggests (Byrne, in preparation),

"It would seem that a potentially fruitful approach to theory building in the science of personality is the construction of small theories dealing with specific and limited behavioral events rather than a general theory of personality. Personality theories should be limited in scope but predictively powerful. As tight and accurate theories are built in areas such as achievment, authoritarianism, and self-consistency, the finding of still more general relationships will tie them together into increasingly inclusive overall theories. This, however, is an end-point rather than a starting point. First, we need to devise psychological laws analogous to those of the pendulum, of gases, of the lever, of the effect of gravity on momentum, etc. As satisfying and useful as it would be to have a complete theory of behavior immediately available, a more limited and more attainable goal should be sought for the present."

Hopefully, the comments in this concluding section will be viewed as a blueprint for future progress in work on repression-sensitization and not as a criticism of any of the work which has been reported. On the contrary, it seems that a great many investigators in a wide variety of locations are engaged in worthwhile and exciting research on this dimension. It is clearly possible to continue these research efforts in such a way as to extend our knowledge about one aspect of personality. And, oddly enough, if we do our work well, it is inevitable that the knowledge will have relevance to all aspects of personality.

References

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., and Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.

Altrocchi, J. Interpersonal perceptions of repressers and sensitizers and component analysis of assumed dissimilarity scores. *J. abnorm. soc. Psychol.*, 1961, **62**, 528-534.

Altrocchi, J., and Perlitsch, Hilda D. Ego control patterns and attribution of hostility. *Psychol. Rep.*, 1963, **12**, 811-818.

Altrocchi, J., Parsons, O. A., and Dickoff, Hilda. Changes in self-ideal discrepancy in repressors and sensitizers. *J. abnorm. soc. Psychol.*, 1960, **61**, 67-72.

Altrocchi, J., Shrauger, S., and McLeod, Mary A. Attribution of hostility to self and others by expressors, sensitizers, and repressors. *J. clin. Psychol.*, 1964, **20**, 233.

Berg, I. A. Deviant responses and deviant people: the formulation of the Deviation Hypothesis. *J. counsel. Psychol.*, 1957, **4**, 154-161.

Berg, I. A. The unimportance of test item content. In B. M. Bass and I. A. Berg (Eds.), *Objective approaches to personality assessment.* Princeton, New Jersey: Van Nostrand, 1959. Pp. 83-99.

Blaylock, Barbara A. H. Repression-sensitization, word association responses, and incidental recall. Unpublished master's thesis, Univer. of Texas, Austin, Texas, 1963.

Block, J., and Thomas, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254-259.

Brown, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

Bruner, J. S., and Goodman, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33-44.

Bruner, J. S., and Postman, L. Tension and tension release as organizing factors in perception. *J. Pers.*, 1947, **15**, 300-308. (a)

Bruner, J. S., and Postman, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69-77. (b)

Byrne, D. The Repression-Sensitization Scale: Rationale, reliability, and validity. *J. Pers.*, 1961, **29**, 334-349.

Byrne, D. Assessing personality variables and their alteration. In P. Worchel and D. Byrne (Eds.), *Personality change.* New York: Wiley, 1964. Pp. 38-68.

Byrne, D. Child-rearing antecedents of repression-sensitization. *Child Develpm.*, in press.

Byrne, D. *Personality: An approach to the science of human behavior.* Englewood Cliffs, New Jersey: Prentice-Hall, in preparation.

Byrne, D., and Bounds, C. The reversal of F Scale items. *Psychol. Rep.*, 1964, **14**, 216.

Byrne, D., and Goldberg, June. Repression-sensitization, authoritarianism, dogmatism, and political attitudes. Unpublished manuscript, 1962.

Byrne, D., and Holcomb, Joan. The reliability of a response measure: Differential recognition-threshold scores. *Psychol. Bull.*, 1962, **59**, 70-73.

Byrne, D., and Sheffield, J. Response to sexually arousing stimuli as a function of repressing and sensitizing defenses. *J. abnorm. Psychol.*, in press.

Byrne, D., Terrill, J., and McReynolds, P. Incongruency as a predictor of response to humor. *J. abnorm. soc. Psychol.*, 1961, **62**, 435-438.

Byrne, D., Barry, J., and Nelson, D. Relation of the revised Repression-Sensitization Scale to measures of self-description. *Psychol. Rep.*, 1963, **13**, 323-334.

Byrne, D., Golightly, Carole, and Sheffield, J. The Repression-Sensitization Scale as a measure of adjustment: Relationship with the CPI. *J. consult. Psychol.*, in press.

Carlson, V. R. Individual differences in the recall of word-association-test words. *J. Pers.*, 1954, **23**, 77-87.

Carpenter, B., Wiener, M., and Carpenter, Janeth T. Predictability of perceptual defense behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 380-383.

Chodorkoff, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, **49**, 508-512.

Couch, A., and Keniston, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151-174.

Crowne, D. P., and Marlowe, D. A. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, **24**, 349-354.

Davison, L. A. Adaptation to a threatening stimulus. Unpublished doctoral dissertation, Univer. of California, Berkeley, California, 1963.

Dulany, D. E., Jr. Avoidance learning of perceptual defense and vigilance. *J. abnorm. soc. Psychol.*, 1957, **55**, 333-338.

Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90-93.

Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

Edwards, A. L. Social desirability and personality test construction. In B. M. Bass and I. A. Berg (Eds.), *Objective approaches to personality assessment.* Princeton, New Jersey: Van Nostrand, 1959. Pp. 100-118.

Endler, N. S. Personal communication, 1963.

Eriksen, C. W. Perceptual defense as a function of unacceptable needs. Unpublished doctoral dissertation, Stanford Univer., Stanford, California, 1950.

Eriksen, C. W. Some implications for TAT interpretation arising from need and perception experiments. *J. Pers.*, 1951, **19**, 282-288.

Eriksen, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230-235. (a)

Eriksen, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, **44**, 442-446. (b)

Eriksen, C. W. Psychological defenses and "ego-strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45-50.

Eriksen, C. W., and Browne, C. T. An experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, **52**, 224-230.

Eriksen, C. W., and Davids, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135-137.

Eriksen, C. W., and Kuethe, J. L. Avoidance conditioning of verbal behavior without awareness: A paradigm of repression. *J. abnorm. soc. Psychol.*, 1956, **53**, 203-209.

Eriksen, C. W., and Lazarus, R. S. Perceptual defense and projective tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 302-308.

Eriksen, C. W., Kuethe, J. L., and Sullivan, D. F. Some personality correlates of learning without awareness. *J. Pers.*, 1958, **26**, 216-228.

Farber, I. E. A framework for the study of personality as a behavioral science. In P. Worchel and D. Byrne (Eds.), *Personality change.* New York: Wiley, 1964. Pp. 3-37.

Festinger, L. *A theory of cognitive dissonance.* Stanford: Stanford Univer. Press, 1957.

Friedman, S. T. The relationship between attitudes toward children and authoritarianism. Unpublished manuscript, 1963.

Gordon, J. E. Interpersonal predictions of repressors and sensitizers. *J. Pers.*, 1957, **25**, 686-698.

Gordon, J. E. The stability of the assumed similarity response set in repressors and sensitizers. *J. Pers.*, 1959, **27**, 362-373.

Gossett, J. T. An experimental demonstration of Freudian repression proper. Unpublished doctoral dissertation, Univer. of Arkansas, Fayetteville, Arkansas, 1964.

Gough, H. G. *California Psychological Inventory manual.* Palo Alto: Consulting Psychologists Press, 1957.

Grigg, A. E., and Thorpe, J. S. Deviant responses in college adjustment clients: A test of Berg's Deviation Hypothesis. *J. consult. Psychol.*, 1960, **24**, 92-94.

Gynther, M. D. Personal communication, 1963.

Hanson, Judith E. Personal communication, 1963.

Hathaway, S. R., and McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. ed.) New York: Psychological Corporation, 1951.

Hereford, C. F. *Changing parental attitudes through group discussion*. Austin: Univer. of Texas Press, 1963.

Holt, R. R. Individuality and generalization in the psychology of personality. *J. Pers.*, 1962, **30**, 377-404.

Joy, V. L. Repression-sensitization, personality, and interpersonal behavior. Unpublished doctoral dissertation, Univer. of Texas, Austin, Texas, 1963. (a)

Joy, V. L. Repression-sensitization and interpersonal behavior. Paper read at Amer. Psychol. Ass., Philadelphia, August, 1963. (b)

Kissin, B., Gottesfeld, H., and Dickes, R. Inhibition and tachistoscopic thresholds for sexually charged words. *J. Psychol.*, 1957, **43**, 333-339.

Kogan, N. Authoritarianism and repression. *J. abnorm. soc. Psychol.*, 1956, **53**, 34-37.

Kurland, S. H. The lack of generality in defense mechanisms as indicated in auditory perception. *J. abnorm. soc. Psychol.*, 1954, **49**, 173-177.

Lazarus, R. S., and Alfert, Elizabeth. The short circuiting of threat by experimentally altering cognitive appraisal. *J. abnorm. soc. Psychol.*, in press.

Lazarus, R. S., and Longo, N. The consistency of psychological defense against threat. *J. abnorm. soc. Psychol.*, 1953, **48**, 495-499.

Lazarus, R. S., Eriksen, C. W., and Fonda, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471-482.

Lazarus, R. S., Speisman, J. C., Mordkoff, A. M., and Davison, L. A. A laboratory study of psychological stress produced by a motion picture film. *Psychol. Monogr.*, in press.

Liberty, P. G., Jr., Lunneborg, C. E., and Atkinson, G. C. Perceptual defense, dissimulation, and response styles. *J. consult. Psychol.*, in press.

Lomont, J. F. Personal communication, 1963.

Lomont, J. F. The repression-sensitization dimension in relation to anxiety responses. Unpublished manuscript, 1964.

Lucky, A. W., and Grigg, A. E. Repression-sensitization as a variable in deviant responding. *J. clin. Psychol.*, 1964, **20**, 92-93.

McDonald, R. L. Ego control patterns and attribution of hostility to self and others. *J. abnorm. soc. Psychol.*, in press.

McDonald, R. L., and Parham, K. J. Relation of emotional changes during pregnancy to obstetric complications in unmarried primigravidae. *Amer. J. Obst. Gynec.*, in press.

McDonald, R. L., Gyther, M. D., and Christakos, A. C. Relations between maternal anxiety and obstetric complications. *Psychosom. Med.*, 1963, **25**, 357-363.

MacKinnon, D. W., and Dukes, W. F. Repression. In L. Postman (Ed.), *Psychology in the making*. New York: Knopf, 1962. Pp. 662-744.

McReynolds, P. A restricted conceptualization of human anxiety and motivation. *Psychol. Rep.*, 1956, **2**, 293-312.

McReynolds, P. Anxiety as related to incongruencies between values and feelings. *Psychol. Rec.*, 1958, **8**, 57-66.

McReynolds, P., and Ullmann, L. P. Differential recall of pleasant and unpleasant words as a function of anxiety. *J. clin. Psychol.*, 1964, **20**, 79-80.

Mathews, Anne, and Wertheimer, M. A "pure" measure of perceptual defense uncontaminated by response suppression. *J. abnorm. soc. Psychol.*, 1958, **57**, 373-376.

Megargee, E. I. Undercontrol and overcontrol in assaultive and homicidal adolescents. Unpublished doctoral dissertation, Univer. of California, Berkeley, California, 1964.

Megargee, E. I., and Mendelsohn, G. A. The assessment of the chronically overcontrolled assaultive person. Unpublished manuscript, 1964.

Nelson, S. E. Psychosexual conflicts and defenses in visual perception. *J. abnorm. soc. Psychol.*, 1955, **51**, 427-433.

Nowlis, V., and Nowlis, Helen H. The description and analysis of mood. *Ann. N.Y. Acad. Sci.*, 1956, **65**, 345-355.

O'Connell, W., and Peterson, Penny. Humor and repression. *J. existential Psychiat.*, 1964, **4**, 309-316.

Page, H. A., and Markowitz, Gloria. The relationship of defensiveness to rating scale bias. *J. Psychol.*, 1955, **40**, 431-435.

Perloe, S. I. Inhibition as a determinant of perceptual defense. *Percept. mot. Skills*, 1960, **11**, 59-66.

Pomeranz, D. M. The repression-sensitization dimension and reactions to stress. Unpublished doctoral dissertation, Univer. of Rochester, New York, 1963.

Postman, L., Bruner, J. S., and McGinnies, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142-154.

Rogers, C. R. A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science*, Vol. III: *Formulations of the person and the social context*. New York: McGraw-Hill, 1959. Pp. 184-256.

Rokeach, M. *The open and closed mind.* New York: Basic Books, 1960.

Schaefer, E. S., and Bell, R. Q. Development of a parental attitude research instrument. *Child Develpm.*, 1958, **29**, 339-361.

Sermat, V. Personal communication, 1962.

Shannon, D. T. The effects of ego-defensive reactions on reported perceptual recognition. Unpublished doctoral dissertation, Stanford Univer., Stanford, California, 1955.

Shannon, D. T. Clinical patterns of defense as revealed in visual recognition thresholds. *J. abnorm. soc. Psychol.*, 1962, **64**, 370-377.

Shrader, W. K. Size-distance perception of human images as a function of self-ideal discrepancy and defensive mode. Unpublished doctoral dissertation, Univer. of Massachusetts, Amherst, Massachusetts, 1963.

Silber, L. D., and Baxter, J. C. Personal communication, 1963.

Solomon, R. L., and Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195-202.

Speisman, J. C., Lazarus, R. S., Mordkoff, A. M., and Davison, L. A. The experimental reduction of stress based on ego defense theory. *J. abnorm. soc. Psychol.*, in press.

Tart, C. T. Frequency of dream recall and some personality measures. *J. consult. Psychol.*, 1962, **26**, 467-470.

Tempone, V. J. Differential thresholds of repressers and sensitizers as a function of a success and failure experience. Unpublished doctoral dissertation, Univer. of Texas, Austin, Texas, 1962.

Tempone, V. J. Construct validity studies of Byrne's Repression-Sensitization Scale. Paper read at Amer. Psychol. Ass., Philadelphia, August, 1963.

Truax, C. B. The repression response to implied failure as a function of the hysteria-psychasthenia index. *J. abnorm. soc. Psychol.*, 1957, **55**, 188-193.

Turk, H. Norms, persons, and sentiments. *Sociometry*, 1963, **26**, 163-177.

Ullmann, L. P. Clinical correlates of facilitation and inhibition of response to emotional stimuli. *J. proj. Tech.*, 1958, **22**, 341-347.

Ullmann, L. P. An empirically derived MMPI scale which measures facilitation-inhibition of recognition of threatening stimuli. *J. clin. Psychol.*, 1962, **18**, 127-132.

Ullmann, L. P., and Lim, D. T. Case history material as a source of the identification of patterns of response to emotional stimuli in a study of humor. *J. consult. Psychol.*, 1962, **26**, 221-225.

Ullmann, L. P., and McReynolds, P. Differential perceptual recognition in psychiatric patients: Empirical findings and theoretical formulation. Paper read at Amer. Psychol. Ass., Philadelphia, August, 1963.

Ullmann, L. P., Weiss, R. L., and Krasner, L. The effect of verbal conditioning of emotional words on recognition of threatening stimuli. *J. clin. Psychol.*, 1963, **19**, 182-183.

Weinberg, N. H. Word association style, field dependence, and related personality variables. Unpublished manuscript, 1963.

Wiggins, J. S. Convergences among stylistic response measures. *Educ. psychol. Measmt.*, in press.

Winder, C. L., and Rau, Lucy R. Parental attitudes associated with social deviance in preadolescent boys. *J. abnorm. soc. Psychol.*, 1962, **64**, 418-424.

Worchel, P. Adaptability screening of flying personnel: Development of a self-concept inventory for predicting maladjustment. SAM, USAF, Randolph AFB, Texas, No. 56-62, 1957.

INTERPERSONAL PERCEPTION

Victor B. Cline

UNIVERSITY OF UTAH, SALT LAKE CITY, UTAH

I. Introduction

Appraising, understanding, and judging others underlies nearly all human intercourse. Most kinds of mental and emotional illness involve fairly serious distortions or breakdowns in communication as well as in interpreting and evaluating social and interactional cues. The evaluation and assessment of others by all men is such a consistent and on-going process that it operates almost automatically at times, commonly at non-verbal and sometimes at almost unconscious levels. It can be a highly rational or even formal process. As Asch has put it, "To take our place with others we must perceive each other's existence and reach a measure of comprehension of one another's needs, emotions, and thoughts."

In a great variety of formal occupations, positions of leadership, and marital, parental, and psychotherapeutic relationships, as well as in informal social units, effective functioning would appear to be critically related to and dependent upon our ability to perceive subtleties and nuances in the behavior of others. We must anticipate their reactions to different decisions and discern those inner states and convictions which, as Gough (1955) has put it, ". . . although determining choice and behavior are often concealed and/or denied in overt expression."

The judging process has been referred to by many names including empathy, insight, clinical intuition, identification, diagnostic competence, understanding, social sensitivity, social acuity, social perception, interpersonal perception (or perceptiveness), person perception, and person cognition. All have slightly varying meanings. For example, "person perception" implies that the judge or perceiver exists apart and outside the world of the person being judged. Such is the case when one uses still photographs of an "other" as the primary source of cues from which judgments are made. If, however, the judging situation involves a double interaction between judge and "other," as usually occurs in real life, it might be more accurately referred to as "interpersonal perception." Tagiuri and Petrullo (1958) have commented, "Through his own presence and behavior in the perceptual situation of the 'other,' the perceiver (judge) may alter the perceptual characteristics of the person whose state he is trying to judge. This is, of course, eminently different from the way in which a rock is a source of cues for a perceiver." Despite differences such as these in the conditions and nature of the judging task, this general area of investigation continues to receive increasing attention from many investigators despite a plethora of methodological complexities.

II. Standardized Judging Tests

When one surveys the literature and such references as "Buros Mental Measurements Year Book" for signs of carefully validated published social sensitivity, empathy, or judging tests, the cupboard is bare. The only tests of note reported in Buros (1959) are Kerr and Speroff's Empathy Test (1951) and Kerr's Primary Empathic Abilities procedure (1957). These, however, tend to be somewhat misnamed in that they attempt to measure sensitivity to the "generalized other"[1] by requiring such things as ranking the circulation of 15 magazines, the popularity

---

[1] "Sensitivity to the generalized other" is a term introduced by Bronfenbrenner et al. (1958) and is defined as an awareness of the social norm or of the typical response of a large class or group. Example: predicting the results of a public opinion poll, or a typical response of a special class of people—say, college students. This is also sometimes called "stereotype accuracy."

of 15 types of music, and so forth. The evidence for validity for these tests is so slight as to eliminate them from serious consideration as genuine measures of empathy.

Gough (1953) developed an Opinion Prediction Scale which seems to measure accuracy of knowledge or stereotype about people in general, or, in other words, sensitivity to the "generalized other." Judges are required to predict the percentage of "true" responses to a number of MMPI-like items for "people in general" and for male prison inmates using a multiple choice response format. The options are 0–25%, 26–50%, 51–75%, and 76–100%. While this is not a measure of judging ability for specific individual persons (often referred to in the literature as "social objects" or "others"), Gough (1955) has found it a useful instrument in some studies of "social acuity."

In addition there are some procedures such as Chapin's Social Insight Test (1942) which have not been published but which have been used in occasional studies with the author's consent on an experimental basis. Knowledge of their existence has been circulated informally as well as through occasional references in the psychological literature. Frequently, as in the case of Chapin's test, the results have been equivocal. With Chapin's test several paragraphs of a true-to-life situation are presented and the judge is asked to choose the correct outcome from several alternatives. In some cases this involves predicting what an individual feels or will do. Other situations require predicting what a group will say, or do. In one study (Gough, 1955) correlations of .47, .47, and .39 were found between Chapin's test and the Terman Concept Mastery Test of Intelligence, the Crutchfield Battery of Insight Puzzles, and the Bennett Mechanical Comprehension Test.

The sound films and judging instruments of Cline (1955a) and Cline and Richards (1958, 1963) are further examples of judging procedures which are still "experimental" and available for research purposes only. These include three entirely separate sets of films developed over a decade, with up to 12 separate judging tests per film and up to 30 films to a set.

From Flanagan's Aptitude Classification Test the "Judgment and Comprehension" subtest has some items that might be regarded as tapping judging ability or social sensitivity though Flanagan makes no claims that this subtest measures other than the "ability to understand a situation and determine the proper action to take." The "How Supervise," (Remmers and File, 1948) sometimes used as an executive or foreman selection test, could conceivably be regarded as an interpersonal perception test. However, its validity data limits it to being a "crude device for screening first level supervisory personnel of low educational level."

The Moss *et al.* (1944) "Social Intelligence Test" might be classed as a measure of empathy. However, existing research suggests that it is a somewhat disguised measure of verbal intelligence and not much more.

TABLE I
VARIETIES OF STIMULUS INFORMATION USED IN MAKING JUDGMENTS

1. Photographs
2. Motion pictures (sound or silent)
3. Unobserved or hidden observation of live behavior (such as through one-way mirror)
4. Face-to-face interactions (from brief interviews through long-term associations such as in work, social, neighborhood groups, or marriage)
5. Tape recordings (using only auditory cues)
6. Test materials and scores (IQ, MMPI, Rorschach protocols, etc.)
7. Written material (biography, autobiography, interview typescripts, descriptive paragraphs, etc.)
8. Personal productions (handwriting, drawings, etc.)
9. Past experience with people in a culture (where the judge is predicting local, regional, or national group norms, such as estimating percentage of population responding "yes" to a particular personality test item)
10. Any combination of the above

TABLE II
TYPES OF JUDGING INSTRUMENTS

1. Trait-rating procedures (adjective checklists, Likert-type rating scales, etc.)
2. Postdicting real life behavior (usually using a true-false or multiple choice format)
3. Postdicting responses to specific test items (such as an MMPI, Strong Vocational Interest Blank, etc.)
4. Postdicting scores on achievement, intelligence, and personality tests (MMPI, IQ, etc.)
5. Postdicting theoretical constructs (such as a psychiatric diagnosis)
6. Writing free global descriptions of social object being judged
7. Using matching techniques (where 2 to x stimulus persons are matched on 2 to x sets of data). Thus one might read personality descriptions of 10 people and try to match up 10 Rorschach protocols to these)
8. Ranking procedures (2 to x stimulus persons are rank ordered on degree of possession of trait or variable)
9. Forced-choice procedures. These are infinite in variety, but for example, might involve choosing which of two statements a pair of "others" agrees or disagrees with
10. Q sort (Here a group of persons can be Q sorted with regard to one personality trait or one person may have 50 variables Q sorted from "most like" to "least like" him)
11. Any combination of above

Most researchers have tended in the past to develop their own judging tests in the absence of any valid procedures available through test publishers. These are usually simple in nature and tailor-made to the

needs of the study. They involve such things as predicting ratings on traits, responses to personality test items, or the like.

Why the conflicting findings in the literature? Analysis of the hundreds of interpersonal perception studies published in the past two decades reveals that the kinds of stimulus information given to judges vary very greatly in type, quality, and quantity. Table I lists a few examples of these. In addition, the choice of judging instrument or procedure can vary from simple rating instruments to elaborate Q sorts, as demonstrated in Table II. When we observe the varieties of sources of criterion information (Table III), and the different possible scoring procedures

TABLE III
SOURCES OF CRITERION INFORMATION

1. (a) Self-provided information. This is where one uses the self-ratings, self-Q sorts, test scores, or other information coming from the person one is judging, as the criterion.

   (b) Group responses to ratings, variables, personality test items, etc. are predicted by judge. This becomes a measure of stereotype accuracy.
2. Associates (peers, subordinates, spouse, teacher, boss, acquaintances, etc.) provide criteria in form of ratings, evaluations, Q sorts, or other types of information about the person being judged
3. Experts (where psychologists, social scientists, therapists, personnel managers, or other "experts" provide ratings, diagnoses, Q sorts, etc.)
4. Combinations of above

available (Table IV), it is easy to understand why searches of the "judging literature" reveal so many apparently contradictory findings. We are dealing with an area of such extreme complexity that experimental methods to determine judging accuracy are seldom identical, and generally varied. There does not exist as yet a single, adequate, demonstrated test of judging accuracy for either global- or component-type measures, though some valiant efforts have been made to produce such an instrument or instruments.

Writers who have argued that there is no general ability to judge others for empathy, intuition, or a general trait, etc. have invariably cited the fact that the findings of judging studies show considerable inconsistency. While this is a legitimate argument, it could be applied to many other areas of psychological research. The reasons for conflicting findings are complex, and cannot be referred to single simple factors.

III. Cronbach's Components and the End of Naïve Empiricism

In 1955, Ronald Taft reviewed the psychological literature in the area of interpersonal perception. He appraised and evaluated several score of the most significant studies to that date. This was, at the time, a definitive and major assessment of the literature. However, later in the

TABLE IV
Scoring Procedures

1. Number of correct predictions (where multiple-choice, accuracy of diagnosis, rank order, true-false, forced-choice, matching, or paired-comparison-type formats are used)

2. Difference score or "D^2 statistic," where rating scales are used; in its simplest form the predicted rating is subtracted from the criterion rating on a "1 to 5 scale" to give a D or difference score. These can be summed over all traits or they can be first squared and then summed. This can also be broken down mathematically into a number of components (see Cronbach, 1955; Cline and Richards, 1960, 1961b, 1962)

3. Correlation statistic (where one correlates a number of predictions made by one judge against criteria or correlates one variable for a number of judges against criteria. Partial correlation technique would be one form of this

4. Quantified evaluations of "free global written description or responses," usually done by experts

5. Special scoring procedures. These can be infinite in variety but two examples from the literature will be cited

    (a) Crow (1954) has developed a "Random-Comparison" method for determining Differential Accuracy which differs from Cronbach's method also discussed in that it is distribution-free, or nonparametric. Whenever the judges make estimations about several "others," a distribution of comparisons is generated by comparing every estimation with every "other's" score, ignoring signs. Thus if eight judges each made an estimation for each of ten "others," the comparison procedure would yield 800 difference scores. The difference score obtained by comparison of the judge's estimation with the score of the "other" for whom the estimation was intended can be compared with the median of the random comparison distribution. If this difference score is more accurate (lower) than the median of the random comparisons, then the judge has obtained Differential Accuracy with that "other." The final score is the percentage of "others" upon which the judge was more accurate than the median of the random comparison distribution

    (b) Bronfenbrenner et al. (1958) developed the "Method of Differential Comparison" as a measure of judging accuracy. They believed this to be independent of the judge's similarity to the person judged and sensitivity to the generalized "other" (Stereotype Accuracy). Any tendency to "assume similarity" could not affect this score. It was obtained by computing the correlation on a within-item basis; that is, by expressing each particular judging estimate as a deviation (x_i) from the mean of the estimates made by the judge for all "others" on that item. Each criterion rating was expressed as a deviation (y_i) from the mean rating given by all "others" for that item. Computation of the simple correlation ($r_{x_i y_i}$) between x_i and y_i over all items and "others" for the particular judge provides an index of the judge's ability to recognize individual differences among subjects in their response to each item. It is also an index of the judge's ability to estimate correctly the *relative position* of self-ratings made by the various "others" on a particular item

6. Nonaccuracy measures. These include such things as Elevation, Differential Elevation, variance of ratings, etc. which do not measure accuracy of judgment per se but may be quite important in interpreting and understanding the accuracy scores obtained

year several papers were published (Cronbach, 1955; Gage and Cronbach, 1955; Campbell, 1955; Hastorf, Bender, and Weintraub, 1955) that seriously challenged nearly all previous research in the area of judging accuracy. Using logic, mathematics, and analyses of studies, it was demonstrated that most judging experiments in the literature had serious flaws which rendered them largely uninterpretable. There had been earlier critical studies (Cohen, 1953; Cronbach and Gleser, 1953; Hastorf and Bender, 1952; Leytham, 1951), and others have appeared since (Crow, 1957; Crow and Hammond, 1957; Gage, Leavitt and Stone, 1956). However, Cronbach's (1955) paper provided a major methodological critique. Using the simple D^2 statistic (i.e., the arithmetical difference between judgment and criterion ratings, squared), he found that this apparently straightforward measure of judging accuracy contained a Pandora's box of "components, artifacts, and methodological problems" which would vitiate many previous research results. Cronbach demonstrated that "difference scores" on trait ratings could be analyzed mathematically into four component parts some of which could be reduced further into variance and correlation parts.

A. Cronbach's Components

1. Elevation (E)

This component measures the difference between the (a) grand mean of an *individual judge's* predictions for all "others" judged on all items and (b) the grand criterion mean. This tells us primarily something about the way in which a judge uses the rating scale and does not deal with accuracy per se.

2. Differential Elevation (DE)

This measures the variance of the individual judge's ratings. It measures the extent to which the judge can predict the deviation from the grand mean of the mean of individual "others" taken over all traits. This score can be broken down into a correlation term (which represents the judge's ability to judge *which* "other" rates highest on the elevation scale) and a variance term which measures the extent to which judged differences between "others" are large or small.

3. Stereotype Accuracy (SA)

This is a measure of what Bronfenbrenner has called "Sensitivity to the Generalized Other," or the degree to which a judge can predict how the whole N of "others" responded to a set of items. The predicted item means averaged across persons are compared with the actual cri-

terion item means across persons. This provides a correlation term which gives the relationship between a given set of predicted item means and the actual criterion means (of all "others" judged) and a variance term which is an index of the variability and/or complexity of the judges' stereotype.

4. Differential Accuracy (DA)

This measures the ability of the judges to predict the differences between "others" on each trait or item considered separately. To put it another way, this measures the difference between the scores for "others" on individual items in the judging matrix *versus* the criterion matrix where in each case the "other's" score is the deviation from both his own mean and the item mean. Thus this can be derived mathematically by subtracting E, DE, and SA from the trait rating total score. It also has a correlation and variance term. This is considered a "pure" accuracy measure by some, even though under some circumstances it also has flaws.

Cronbach (1955) concluded that, "Social perception research has been dominated by simple operationally defined measures (of judging accuracy). Analysis has shown that any such measure may combine and thereby conceal important variables or may depend heavily on unwanted components. Only by a careful subdivision of global measures can an investigator hope to know what he is dealing with." Later (1958) he commented, "Investigators have been attracted to this area by their interest in such constructs as transference, identification, projection, empathy, and insight. We must question whether these terms should be employed in future interpretations of social perception experiments. If a behavior which looks like 'projection' can arise out of many different processes, there is little point in trying to formulate hypotheses using the concept 'projection.' Instead, theories of perceptual response should take into account the traits being perceived, the constant tendencies in this perceiver (judge) with respect to those traits, and finally the effect of the particular 'other' as a social stimulus to this perceiver. Worthwhile hypotheses of this order cannot be formulated until data from exploratory studies have been treated *analytically* (i.e., mathematically broken down into component scores) . . . a single global index (of judging accuracy) can now be replaced with a series of components of the perceptual relation and a separate analysis can be made for each. Even the simplest examples make it evident that insight, assumed similarity, and the other global phenomena are the shadow not the substance of social perception."

In recent research by Cline and Richards (1960, 1962), and Richards and Cline (1963) it was demonstrated that the system used in assigning

numerical scores to the characteristics (traits) included in the judging matrices could produce artifacts tending to reduce or, in extreme cases, to eliminate Stereotype Accuracy and Differential Elevation correlation terms. Cronbach's Differential Accuracy in certain circumstances is thus made difficult or impossible to interpret. These artifacts were seen as particularly objectionable in the case of Differential Elevation and Differential Accuracy. To overcome these difficulties, Cline and Richards (1962) and Richards and Cline (1963) proposed a new measure, Interpersonal Accuracy (IA), which appears to be the linear combination of Differential Elevation and Differential Accuracy but is free of the aforementioned artifactual effects of the scoring system. This new component is discussed in detail later in this paper.

Other investigators have become increasingly chary of naïve empiricism in interpersonal perception research with the omnipresent danger of misinterpretation of results. A good example of this is Leytham's (1951) review of Taft's (1950) study of the personality correlates of judging ability. In Taft's research, using 40 graduate students, he studied judging ability by requiring each subject to rate a number of his fellow assessees, with whom he was familiar. Six personality variables were rated on a conventional five-point scale. Each judge also predicted how his peers would rate him. The criterion was the consensus of experts (research staff) and peer ratings. The more accurately the judge rated others and predicted self ratings, the better judge of personality he was assumed to be. The next step was to find test and rating correlates of judging ability as thus defined. Leytham later pointed out a statistical artifact, i.e., that cautious judges (who made many "three" or average ratings) were given a major advantage over those who tended to use the more extreme ratings. If judging accuracy were computed by the sum of the differences from the criteria consensus ratings then the rater who gave *nothing but* "three" or average rating would be easily the "most accurate" judge, because he could never obtain a discrepancy or error score greater than two (i.e., $5 - 3 = 2$ or $3 - 1 = 2$). Since the criteria consensus ratings (a summary mean score) tended to regress toward the mean of three, this amplified the spurious advantage already given to the cautious judges. Thus when Taft correlated his judging accuracy scores with other outside measures, he found a relationship between them and a "response set" to judge cautiously, but not necessarily "accurately."

Cronbach (1955) has commented on this point: "If two diagnosticians can each judge some trait with the correlational validity of .40, the one who differentiates strongly (i.e., makes extreme statements) will make far more serious absolute errors than the one who differentiates moderately. Indeed, the person who makes extreme differentiations based on a validity

of .40 may have larger errors than a judge who has zero correlational validity but gives the same estimate for everyone."

In another example, Dymond, Hughes, and Raabe (1952) had sixth grade pupils rank each other sociometrically and also predict what sort of ratings they (the raters) received from the others. They found that accuracy of judgment or estimation correlated .50 with sociometric acceptance. However, the students generally tended to predict that they would be highly accepted. This meant that those students who, in fact, were highly accepted automatically obtained higher accuracy scores. This fact alone is sufficient to account for the apparent relationship between judging accuracy and sociometric acceptance.

In a study by Nagle (1954) 14 supervisors were each asked to predict the number of their own subordinates responding in a favorable direction to some morale items having to do with their attitudes toward the supervisor. Accuracy of prediction was obtained for each supervisor by summing, over the items, the *differences* between the supervisors' predictions and the in-fact responses of the subordinates. For each one of the supervisors' groups a summary morale score (based on subordinate attitudes toward the supervisor) was obtained from this same instrument. A correlation of .90 was found between supervisors' empathic accuracy and the morale index. Nagle thus concluded that empathy, or judging accuracy, was related to "morale," as defined in this study, and his data apparently supported this conclusion. However, a more parsimonious explanation would be that most supervisors predicted highly favorable attitudes toward themselves by the men in their group. Thus the supervisor whose group morale (as indicated by the favorable attitudes shown toward him by his subordinates) was highest would have to receive the highest empathy or judging score. Spuriously high scores on empathy measures will be made by the truly popular subject when his empathy is tested by ability to assess his own popularity; a common human response set toward rating oneself as popular will guarantee this even in the most imperceptive of popular people.

B. The RS-AS-ACC Problem

Various investigators have been concerned with the RS-AS-ACC problem. In this case,

a = the judge's self-description on a trait variable,
b = the "other's" (person being judged) self-description,
c = the judge's prediction of b.

Real Similarity (RS) is where $a = b$, *Assumed Similarity* (AS) is where $a = c$, and *Accuracy* (ACC) is where $b = c$. Many social perception

studies have gone awry because of the lack of understanding that while any two of these relationships are independent, the third remaining relationship is directly and mathematically linked to and a resultant of the other two. Hatch (1962) has commented that, "Where each of a number of judges possesses high RS with their others (persons being judged) their accuracies will correlate perfectly with the degree to which the judges assume similarity with their others. A typical situation is one in which each of the judges predicts with a high degree of assumed similarity. The accuracy score then varies directly with the amount of real similarity existing between the judges and their respective others. This latter relationship, real similarity, may best be viewed as a fixed constant which exists between a judge and a specified other, as a function of the sampling. With the high degree of assumed similarity often found in empathy studies (especially where judges predict for some variable allowing self-flattering predictions), the resulting accuracy scores bear a highly significant relationship to this original constant, RS. Such accuracy scores certainly do not represent novel emergents reflecting on empathic sensitivity of some kind. They are simply the logical consequences of RS-AS-ACC linkages and are impossible to interpret."

An example of an "RS-AS-ACC study" is Smith's (1957) study of rapport and social perception in mother-adolescent relationships. Some 15 different judging scores were derived from the four types of data listed below.

(a) Daughters took Kuder Preference Record for themselves;
(b) Mothers predicted their daughter's responses on this Kuder test;
(c) Mothers predicted typical adolescent responses to the Kuder test;
(d) Mothers took Kuder for themselves.

The following 15 scores were derived from this information.

(1) *Accuracy* (ACC): Number of items on daughter's Kuder which coincide with mother's predictions for her.

(2) *Real Similarity* (RS): Number of items on which mother's self-Kuder and daughter's self-Kuder coincide.

(3) *Real Dissimilarity* (RD): Number of items on which mother's and daughter's self-Kuders differ.

(4) *Assumed Similarity* (AS): Number of items on which mother's self-Kuder and mother's prediction of her daughter's Kuder coincide.

(5) *Assumed Dissimilarity* (AD): Number of items on which mother's self-Kuder and mother's prediction for her daughter differ.

(6) *Warranted Assumed Similarity* (WAS): Number of items on which mother's self-Kuder, daughter's self-Kuder, and mother's prediction for her daughter all coincide.

(7) *Warranted Assumed Dissimilarity* (WAD): Number of items on which mother's self-Kuder does not agree with daughter's self-Kuder, but does agree with mother's prediction for her daughter.

(8) *Unwarranted Assumed Similarity* (UAS): Number of items on which mother's self-Kuder agrees with her predictions for her daughter but not with the daughter's actual self-Kuder responses.

(9) *Unwarranted Assumed Dissimilarity* (UAD): Number of items on which the mother's self-Kuder coincided with her daughter's self-Kuder but not with the mother's predictions of her daughter's Kuder responses.

(10) *Stereotype Accuracy* (SA): Number of items on which mother's prediction of a typical adolescent response on Kuder coincide with her daughter's self-description, or responses, on the Kuder.

(11) *Stereotype Inaccuracy* (SI): Number of items on which mother's prediction of the typical adolescent response differs from her daughter's self-description or responses on the Kuder.

(12) *Assumed Conventionality* (AC): Number of items on which mother's prediction of a typical adolescent response and her prediction of her daughter's Kuder responses coincide.

(13) *Assumed Unconventionality* (AU): Number of items on which mother's predictions of a typical adolescent response and a prediction of her daughter's Kuder response differs.

(14) *Warranted Assumed Conventionality* (WAC): Number of items where mother's prediction of typical adolescent responses, her daughter's self-description, and the mother's prediction of the daughter's Kuder responses all coincide.

(15) *Warranted Assumed Unconventionality* (WAU): Number of items where mother's prediction of typical adolescent responses agrees with her daughter's prediction of her daughter's responses, but does not agree with her daughter's actual self-description on the Kuder.

Smith found low positive correlations between mother-adolescent offspring rapport and judging accuracy scores. However, the evidence suggested that this accuracy was due in part to the tendency of high-rapport judges (mothers) to assume similarity and also possibly to assume conventionality. But because of the methodology of the study no evidence could be presented showing how much of the accuracy variance could be attributed to these artifactual processes.

Cronbach's (1955) previous analysis of the components of judging scores emphasized the "difference score" already described. An alternative is the correlation between predicted scores and criterion scores. These methods may be compared for their respective freedom from artifacts.

Crow (1954) commented that, "The use of the correlation scoring

method defines predictive accuracy as the ability to vary one's predictions as the actual situation varies. The 'difference score' method defines predictive accuracy as the ability to approximate the actual situation. By the 'difference score' method a subject (judge) is penalized for a systematic error in estimation of the magnitude of the actual situation. By the correlation method the subject (judge) is not so penalized. Conversely, a subject (judge) is penalized by the correlation method, if although he has approximated the actual situation, his predictions do not vary concomitantly with the actual scores."

Thus, in the correlation method such response sets as Cronbach's Elevation or those associated with the variability of prediction and response are removed, though the RS-AS-ACC linkages remain. It should be noted that a correlation approach can require a judge to predict for many "others" on a single variable, IQ for example, or it may require him to make predictions for a single "other" on a large number of separate trait variables. In either case methodological complexities remain which can render judging accuracy scores suspect. For example, special problems in interpretation arise in the case where a correlation is calculated between criterion and a judge's predictions of the self-score of each of a number of "others" on a single variable. Here, as Hatch (1962) has pointed out, there is a tendency for the judge to assume similarity (defined as perceiving others as similar to oneself). This must be considered as a response set and this will differ from "other" to "other" apparently as a function of the judge's generalized attitude toward the "other."

The tendency to assume similarity is fairly constant over preferred "others." However, knowledge of the degree of assumed similarity between a judge and a liked "other" provides no basis for predicting the degree of assumed similarity which will be exhibited by the judge for a disliked "other."

Gage and Cronbach (1955) have noted that assumed similarity is highly general over items, and people tend to assume similarity to the same degree throughout a questionnaire despite marked variety in the item content with this tendency being somewhat general for preferred "others." Thus, a judge's attitude and set toward an "other" will influence the AS score.

Hastorf and Bender have tried to clarify and solve the problems associated with ACC, or accuracy, which they call "raw empathy," by subtracting AS from it to give a "refined empathy" score. The major problem here is that they lose important data when they do this because some of the subtracted AS data represent true accuracy of judgment. Gage and Cronbach (1955) in an excellent discussion of this note that

spurious relationships can occur wherever the "other's" self-score correlates with the "outside" variable, which is to be paired in a correlation with the judge's accuracy scores. They suggest that researchers test for the presence of spurious or artifactual relationships between the accuracy and empathy scores on the criterion variables by checking the relationship between the "response to be judged" and these same criterion measures. An example is given to further illustrate this point. Employees are asked to rate their department supervisor and he in turn predicts their ratings on a 1-to-5 point scale. If every supervisor predicts that his group will rate him 5 (very good), as might be expected, then there will be an automatic perfect correlation, 1.00, between the supervisors' "empathic ability" and receiving high ratings as a supervisor.

C. Other Sources of Error in Judging-Accuracy Scores

To give the reader an even more detailed and clearer idea of the kinds of biases and response sets, etc., which can add error variance as well as "chance success" to judging-accuracy scores, a few will be listed below.

(1) *Social desirability bias:* This refers to the tendency for judges usually to predict the most socially desirable response in making judgments and predictions about "others."

(2) *Similarity of the judge to the "other":* This may be in sex, age, personality, religion, occupation, class level, ethnic, social, and cultural background. This in turn may lead the judge to believe that "the other" is like himself, and respond to the items as if this were so. This is very similar to (9) below.

(3) *Acquiescence set:* This refers to a tendency to agree with or predict "Yes" to items rather than to choose the negative or "No" responses.

(4) *Use of stereotype:* Wherein the judge uses an undifferentiated stereotype of "fraternity men," "Jewish women," etc. to predict for specific "others," and only if his "other" fits the stereotype will he achieve accuracy.

(5) *Personal reactions:* Liking or disliking the individual being judged can produce a "halo" effect in rating or judging.

(6) *Making use of an implicit personality theory:* Wherein the judge assumes there is an invariant relationship between trait *a* (observed in the "other") and traits *b, c,* and *d* (not observed but assumed correlated).

(7) *Tendency to make extreme ratings or judgments:* This is a response set wherein the judge overdifferentiates in his ratings of "others."

(8) *The central tendency response set:* Wherein the judge sees most "others" as being very similar or alike and gives mostly middle range or average scores with very little differentiation from "other" to "other."

(9) *To assume similarity or project:* In this case, the judge sees "others" as being very similar to himself (as defined by his self-ratings).

(10) *To assume dissimilarity or project "in reverse":* Wherein one perceives most "others" as being different from oneself.

(11) *Semantic ambiguities in trait terms:* Here the judge interprets the trait name in the judging instrument assuming a meaning other than that intended in its criterion development and use.

Mathematical linkages between predictor and criterion variables (as already discussed), and the use of "global scores" which contain numerous components sometimes working in opposite directions, can also be a further source of error. Several investigators (Cline and Richards, 1960; Hatch, 1962; Crow, 1954; Bronfenbrenner, Harding, and Gallwey, 1958; Grossman, 1963) have attempted to design studies that would meet these objections. However, their researches have met with only partial success.

Grossman (1963) used 5-min movies of adult subjects being interviewed. Three films were of males and three of females. These sound-color movies had been developed previously by Cline and Richards. By showing three sets of film at a time and using a matching method, it was possible for Grossman to control or eliminate the Elevation and Differential Elevation components as discussed in Cronbach's 1955 paper. Grossman then attempted to eliminate Stereotype Accuracy, the third of Cronbach's four components, by the following technique.

Starting with 240 judging items for the six films, two item analyses were undertaken. The first was an internal consistency analysis to choose the most highly discriminating "interpersonal sensitivity" items. The second involved giving a four-part test on *group* sensitivity to the same subjects (judges). This test was believed to be a measure of Stereotype Accuracy. Items retained in the final judging test were highly discriminating for interpersonal sensitivity, but nondiscriminating for group sensitivity or Stereotype Accuracy. The final correlation between the interpersonal sensitivity index and the group sensitivity index was .12, or nonsignificant. The final judging test thus derived was believed to be a relatively pure measure of Differential Accuracy. This index was correlated with 38 personality, demographic, and other types of variables. Grossman found some evidence for generality of judging ability. His cross-sexes correlation when corrected for attenuation was .48. His cross-persons correlation when corrected for attenuation was .60. He also found that the more observant a judge was, the higher his judging score. He obtained a measure of observant accuracy by testing how much the judge could remember about the appearance, actions, and behaviors of the people in the film. He also found that the better judges tended to be more "open-minded" as measured by three subscales, Religious Skepti-

cism, Nonconformity, and Liberalism. He also found low positive correlations between judging ability and intelligence, using ACE Total, Quantitative, and Language scores.

D. HATCH'S FORCED-CHOICE DIFFERENTIAL ACCURACY APPROACH

Hatch (1962) tackled the methodological problems of interpersonal perception research with great care, going to considerable lengths to control bias, response set, and other artifacts. He identified three primary difficulties to be overcome in any research of this kind: (1) solving the RS-AS-ACC problem; (2) controlling for miscellaneous response sets; and (3) developing an adequate "better than chance" test for one's judging-accuracy results. He first proceeded to develop a Forced Choice Differential Accuracy Test which gave a score corresponding to Cronbach's Differential Accuracy correlation term. The judges (30 branch managers of Minnesota Mining and Manufacturing Co.) were required to predict which *one* of two statements, both of a pair of their subordinates *disagreed* on. *Example* (slightly modified):

Mr. Brown

Mr. Jones

Check One

1. My sales manager assigns first priority to problems brought to him by a salesman.

☐

2. My sales manager should consult me more often for my opinion or advice.

☐

A large number of subordinates (318) previously responded either Strongly Agree, Agree, Disagree, or Strongly Disagree to items in a sales attitude questionnaire from which the pairs of men and their responses to items had been taken in developing this empathy test. Hatch found that any two subordinates might have identical responses (IR) or non-identical responses (NIR) to an item. There were three kinds of NIR response subcategories: (A) Maximally Dissimilar Responses (MDR) where one subordinate says "Strongly Agree" *versus* "Strongly Disagree" for the other; (B) Dissimilar Responses (DR) of Agree *versus* Disagree; and (C) Similar Responses (SR) of Agree *versus* Strongly Agree, or Disagree *versus* Strongly Disagree. All three of these were classified as "disagreed on" (or NIR) responses of the supervisor's subordinate pair in the empathy test.

Using this Forced Choice Differential Accuracy technique, Hatch demonstrated empirically that the structure of the forced-choice items did not permit a confounding statistical linkage to exist between RS, AS, and ACC. Thus, the judge might obtain accuracy under any combination of RS and AS conditions. "Accuracy depends only upon the correct identification

of the NIR statement and is uninfluenced by the real similarity and high-low assumed similarity conditions; thus, judges with any combination of RS-AS relationship may be scored as accurate or inaccurate." Hatch arranged to control for four specific types of response sets and set up his experimental design so as to control "unidentified response sets." These four response sets were identified as "response variability," "response intensity," "atypical response pattern of judge," and "atypical response pattern of others." The first ("response variability") will be discussed briefly for illustration purposes. From the responses of salesmen to the original sales attitude questionnaire, it was found that some items were controversial, with much disagreement among respondents. With other items there was very little response variance. If most of the NIR statements had high response variance, it would be possible for a judge knowing only this to obtain a high accuracy score by always picking the most controversial item, as the empathy test required predicting which of two statements the subordinates disagreed on. To control this, statements were matched so that the IR statement variance exceeded the NIR statement variance in exactly one half of the forced-choice items prepared for each judge. Elaborate experimental controls in item selection were also established for the other three response sets mentioned previously.

In an effort to assess the practical validity of his empathy test, Hatch correlated it with (a) a measure of human relations skill (by having company executives nominate branch managers who were most and least successful in interpersonal relations with subordinates), (b) degree of acquaintance (between branch manager and subordinate), and (c) degree of extrapolation or inference required between information input and outtake (prediction). The latter item was assessed by comparing judges' accuracy on each of the three NIR items (SR, DR, and MDR) matched with the IR item. Thus, the greatest extrapolation or inference required of a judge (and also theoretically the most difficult judgment) would be on an IR *versus* SR item, whereas the easiest would be an IR *versus* MDR item. Each manager made predictions on 12 SR items, 12 DR items, and 12 MDR items. In this way the degree of measured discordance in the attitudes of a subordinate pair was varied for each manager, though held constant from manager to manager. The subordinates under each branch manager were also classified into three groups: (a) AP (attitude predictable); these were the subordinates that a branch manager felt most confident in making predictions about and with whom he was most acquainted; (b) AU (attitude unpredictable); these were the subordinates that a branch manager felt least confident in making predictions about and was least acquainted with; and (c) AD (attitude divergent); these

were salesmen at either extreme on a "dissatisfaction index" taken from the sales attitude questionnaire. Each judge predicted on 12 items for each of the three pairs of subordinates, AP, AD, and AU.

Each judge (branch manager) received a different 36-item pair inventory devised individually for him. Because of the highly restrictive conditions set in controlling the four response sets in addition to AP, AU, AD and SR, DR, and MDR conditions, it was necessary to program a "look up and comparison" table using the Univac 1103 computer. Over 14,000 IBM cards and 20 hours of computer time were necessary to list the eligible items in order of priority. Following this, some 40 additional hours were involved in selecting the 10,080 forced-choice items from the computer printout. Hatch concluded that the costs were excessive and that the Forced Choice Differential Accuracy approach, "Is probably not a practical solution to the empathic measurement problem." It was also indicated that if the same computer data were used by another investigator, precisely the same inventories would *not* result and, in fact, Hatch would not be able to repeat the process if he were, himself, to do it again.

The final results indicated that empathy as measured here was not related to human relations skills but was somewhat related to degree of acquaintance between manager (judge) and subordinate as well as degree of inference or extrapolation required. While his good judges as a group were able to predict at better than chance ($P < .05$), Hatch noted that "much but not all of the prediction of the study could be accounted for by chance alone."

Because this study represents a major effort to eliminate the methodological gremlins which plague interpersonal perception research, it is a significant contribution to the literature. Yet the lack of any relationship with human relations skills, as well as the only slightly above chance accuracy scores obtained by judges, is of considerable concern to this reviewer. Serious questions arise about the potential future of an approach where such costly and careful controls yield such modest results.

E. Accuracy versus Process Approaches in Research

In the person perception literature a sharp cleavage is apparent between research which emphasizes accuracy of judgment and that which focuses on process (i.e., how we judge).

A growing number of researchers have abandoned work on the assessment of judging ability as a trait in favor of the view that this exists only in a specific instance with a particular person using a certain instrument.

Tagiuri and Petrullo (1958) have tried to turn the tide in favor of research emphasizing process by the publication of a major collection

of papers contributed predominantly by workers in this group. Tagiuri, in the same volume, has stated the case with considerable vigor: "For a number of reasons, attempts at studying correlates of accuracy have with very few exceptions produced negligible correlations and yielded very little insight into processes. First there is no single satisfactory criterion against which to match the judgments. The criteria used—objective behavior, self ratings by the object person, ratings by the experts, consensual ratings by peers—do not always agree and have very different psychological implications. Second, the disparity of tasks and abilities subsumed under the various operations called measures of accuracy have been glossed over. It is also probable that different judgmental skills may be involved in different situations. In addition, most accuracy scores contain some seven different and not necessarily correlated components. There is furthermore the extreme dependence of results upon judgmental sets and upon the distributions of the variables that are to be judged. Finally, most of the studies are inconclusive because of the lack of representativeness in the design employed. In sum, investigations yield data that are difficult to interpret and impossible to compare. It is the process rather than its achievement that one must investigate if a broad understanding of the phenomenon is to be reached. This point cannot be stressed sufficiently. It is also important to realize that the difficulties encountered in quantitative studies of *accuracy* cannot be eliminated by resorting to careful qualitative approaches. Such phenomena as real and assumed similarity, stereotype and differential accuracy, favorability sets, and artifactual relations, to mention but a few, apply to any kind of inquiry into person perception."

Fiedler (1960) stated that he is not interested in accuracy of judgment per se, but only in what kinds of perceptions, regardless of accuracy, are related to such behavioral outcomes as winning basketball games, accuracy of bomber crews, etc.

However, Crow (1960) has responded to this position by commenting, "Only in recent decades has physical perception been freed from the preoccupation with mediational detail and the stimulus bias that has characterized its history. In person perception research, we must examine cautiously any call to focus on process that turns its back on accuracy. My point is not that we should ignore process nor should we fail to eliminate from our scores those components which invalidate our conclusions. The danger lies in the possibility that we will ignore accuracy and that the components once eliminated will not receive the attention they require. We should study process but we should not become bogged down in the technological details of the machinery by which the organism arrives at interpersonal perceptions. We must keep in mind that the process we are

studying is a process of functional achievement. If we fail to do this in person perception research we may engage in a tedious reinactment of the history of physical perception."

IV. Analysis of Judgments from Motion-Picture Samples

The present author has been conducting a series of research studies in the area of interpersonal perception since 1951. Many have been in the nature of large-scale pilot studies which have frequently yielded negative results and have not been published. However, negative results are frequently as important and necessary to progress as positive findings and both kinds will be reviewed in the following section.

In preliminary consideration of the literature the writer noticed the frequent use of still photographs as stimuli for judgments. While this kind of "limited cue" study has its place, the writer is in agreement with Bruner and Tagiuri (1954), who commented that "A vast literature arose based on judgments of emotions from still pictures and drawings of human faces, when we know that it is rarely that one makes a judgment based upon a frozen millisecond of exposure to a face expressing emotion with all other forms of information lacking—it seems unlikely that conventional analysis of expression carried out with callipers will yield much value and understanding of stimulus properties of faces in a state of emotion. . . ."

The writer and his collaborators were of the opinion that a more dynamic true-to-life type of stimulus was needed. A sound film of an interview (following the work of Estes, 1938; Gage, 1952; Giedt, 1951; Luft, 1949; and others) appeared to be a possible alternative to immobile photographs or drawings of faces. This permits the recording of a variety of verbal and visual cues in situations allowing for repeated observation. Interview behavior appears also to be more typical of the kind of phenomenon upon which judgments are made in real life. In early research conducted at the University of California by the writer, a dilemma presented itself. It was found on an experimental basis that 2 hours proved to be the maximum period for which most judges could tolerate looking at films and fill out judging instruments, at least at one sitting. Thus, one might make a film of an interview lasting 45 minutes to an hour and present one or two of these to a group of judges for evaluation, but the number of "others," (social objects or persons judged) was extremely small and limited. Or as another alternative the experimenter could present a fairly large number of 1-minute filmed interviews, but so limit his judges in the amount of information presented to them that it almost approached the still photographs as a source of data. Thus, in time, it became apparent that the filmed interview could not be a typical, un-

structured, loose, clinical type of interview, but must in as short a time possible compress, capture, and present a wide range of information both objective and factual in content as well as emotional and reactional in nature.

A. CONSTRUCTION OF FILMS

After some 6 months of experimentation, the following procedure was developed. Sound motion pictures in black and white were taken of nine male college students using a hidden camera and microphone. Interviewees initially knew only that they were to report for job interviews in their particular fields of study. The interviews were very highly structured and were held relatively constant for all those participating. Each interview had three phases: (1) A standard opening session in which stock employment questions were asked by a mature and skilled actor posing as a prospective employer; (2) a stress session in which the "employer" became critical and mildly stressful in his questioning; and (3) an abreaction session in which the inteviewer stopped playing the role of the employer, revealed the true experimental intent of the interview, and became quite humble and deferential. He invited comments about how the interviewee had felt in such a situation and what had bothered him the most. It was at this point that the interviewee relaxed greatly and usually became quite verbal and very frank, indicating points of tension, blocking, embarrassment, and his degree of insight into the total situation. This was all recorded by the hidden camera and microphone. Permission was then secured to use the films for research. The four best films of the nine, representing the most diverse personalities, were presented to large populations of judges.

This employment interview which was in part a stress interview, had been suggested by one of the OSS (1948) situational procedures used during World War II. In developing this technique for the present research everything possible was done to compress in 11 minutes a great range of information about the individual being interviewed, both objective and emotional.

Three judging tests were constructed.

1. Behavior Postdiction Test

Here the judge was required to describe or postdict how the interviewee behaved in everyday life. A series of questions was set up with a multiple-choice format. Example:

> 14. When the interviewee (in the film) is in a violent argument usually he:
>
> (a) Becomes very sarcastic,
> (b) Uses profanity and obscene words,

(c) Leaves the room or area,
(d) Strikes his opponent with his fists.

There was only one correct response. The other alternatives were were carefully constructed so as not to include even partially correct statements. A separate and individually tailored Behavior Postdiction Test was constructed for each interviewee. Other items tapped such areas as relations with women, behavior at social gatherings, handling money, etc. The experimenter knew how each interviewee characteristically behaved, and was able to write items tailored specifically to his personality because intensive interviews had been held with him, and his family, close friends, fraternity brothers, and/or fiancee, or wife, where available. Only those behaviors which all agreed upon as being characteristic were used, and the incorrect alternatives were chosen with equal care so as to eliminate even partial or occasional kinds of behaviors. The judging score was the total number of correct predictions made for all four filmed interviews.

2. *Personality Word Card*

Next the judges were required to predict the interviewees' "verbal behavior." This was done in the following manner. The interviewees had on three previous occasions, about a month apart, filled out the Personality Word Card, checking those adjectives which they felt best described themselves. This card was merely a group of 200 descriptive adjectives such as, "frank," "bossy," "sexy," "persistent," "anxious," "aggressive," and so forth. It was found that certain adjectives were "marginal," that is, the S checked them on only one of the three occasions he filled it out for himself. Other adjectives proved to be more central and important in nature, being checked as self descriptive two and/or three times out of three. These were retained in a final key. Thus, each interviewee had a unique pattern of adjectives which represented an important and consistent part of his self perception. Each judge then was required, after seeing the film, to predict which adjectives had been checked by the interviewee in describing himself. Each judge was told to check only about one third of the adjectives in making his predictions on each interviewee because the interviewee keys average one third the total number of adjectives. The judging accuracy score for each judge was based upon the percentage of correct predictions averaged for all four of the test films.

3. *Multiple Choice Sentence Completion Test*

This technique was also used but abandoned because of low reliability. Here the judge attempted to predict the sentence completion

responses, using a multiple-choice format, for each of the interviewees in the films.

B. Preliminary Reliability and Validity

Reliability coefficients for each of the three techniques were computed by employing the Spearman–Brown formula to the correlation between odd and even halves. The Postdiction test yielded a reliability of .56. The Personality Word Card yielded a reliability of .83. The reliability of the Multiple Choice Sentence Completion Test was only .35, and for this reason was discarded as an index of judging ability. None of these were corrected for attenuation, which would have given them higher reliabilities. An equally weighted composite of scores on the first two judging tests (Postdiction Test and Personality Word Card) was obtained for each judge. These two measures intercorrelated $.43 \pm .10$. In the initial experiment five groups of judges viewed these four films and made judgments. After each filmed interview had been shown the projector was stopped and they were asked to make their judgments on the instruments discussed. These Ss were (a) 109 undergraduate college students, (b) 106 professionals (clinical psychologists, psychiatrists, and graduate clinical trainees), (c) 47 adult members of a Protestant church congregation, (d) 43 nursing trainees, and (e) 11 advanced engineering students.

The results may be briefly summarized as follows: All groups of judges performed at very much better than chance success (.001 level of confidence) on both the Behavior Postdiction Test and the Personality Word Card. The question arose, however, "To what extent would the judges have done just as well without seeing the films but merely filling in the tests according to their stereotype of a typical college male?" To answer this question a special control group of 57 undergraduates filled out the forms without seeing any of the films whatsoever, merely using hunches or social stereotypes about college males, in responding to the various questions. Their level of accuracy was termed "psychological chance." When the control and experimental groups of judges were compared in their performance the test indicated significant differences, better than the .001 level of confidence, on both judging instruments, favoring those people who had *seen* the films. The evidence seemed to be fairly conclusive that the judges who saw the films were making fairly accurate predictions or judgments on the basis of a differential analysis and a real evaluation of the personalities of the films rather than from some crude internalized stereotypes of what college males were like.

This study was conducted in 1952 prior to the publication of most of the articles suggesting components, RS-AS-ACC problems, and the

like. However, it should be noted that the D^2 statistic, which later proved so fallible, was not used in this study, even though the possibility of other "error" or artifacts remains.

When the judges of the five subgroups were compared for accuracy a three-level hierarchy was found. Most accurate and proficient were the professionals, that is, the psychiatrists, psychologists, and clinical trainees. At level two were the nursing trainees, and at level three the college students and members of the church congregation as well as the advanced engineering trainees. The differences in accuracy between levels one and three on both the Postdiction Test and Personality Word Card were significant at the .05 level of confidence. The differences between levels one and two, and two and three, however, did not quite reach significance. It was also found that women consistently obtained slightly higher judging scores than men. While these differences never quite reached significance they were consistent in this study as well as most other studies conducted by the author. It was also found that some of the interviewees or "others" in the films appeared generally easier to judge than others. Length of professional psychological or psychiatric experience was related to (a) greater accuracy in using the Personality Word Card, that is, predicting "verbal" behavior, and (b) decreased accuracy in predicting real life social behavior on the Behavior Postdiction Test. This finding was based upon a comparison of the accuracy of predictions made by psychiatrists and psychologists having more than ten years experience with the accuracy of predictions made by those having less than three years experience. The relatively new and less experienced clinicians were a little less accurate in predicting verbal behavior ($P = .10$), but were considerably more accurate in predicting real life behavior ($P = .02$) than were their much more experienced and "older" colleagues. This suggested that there may be danger with increasing experience in becoming more sensitive and aware of verbal behavior but less aware and less in contact with reality with real life social behavior. The top and bottom quartiles in judging ability of the college sample filled out Gough's Adjective Check List describing themselves. The better judges significantly more often characterized themselves on this check list as "sympathetic" and "affectionate." The adjective self checks characterizing the poorest judges included "dissatisfied," "irritable," "awkward," "praising," and "hurried."

With regards to the personality correlates of judging ability, as here defined, results were available for the sample of 109 college undergraduates. They were given a battery of tests which included the MMPI scored for 22 scales, the California Public Opinion Scales (including E, F, and PEC), the Henmon–Nelson Intelligence Test, the Gough Adjective

Check List, the Brunswik Faces Test (wherein one estimates the intelligence, likeability, and so forth of 46 males on the basis of small still photographs of their faces); the Barron–Welch Art Scale, the Group Opinion Estimate (where one attempts to predict what percentage of the general population responds true to each of 30 MMPI items, such as "I pray several times a week"), and the Klein Social Prediction Technique, wherein one is given a transcript of a patient's remarks in a clinical interview that is interrupted at a crucial point, and five possible endings are presented, only one of which is correct (the task here being to predict the correct alternative).

In summary it was found in this study that the most significant correlates of judging ability were (a) absence of ethnocentric-authoritarian attitudes (Fascism and Ethnocentrism correlated −.46 and −.32), (b) superior intellectual ability, (c) lower scores on MMPI scales for Hypochondriasis, Dissimulation, Paranoia, Schizophrenia, Psychopathic Deviate, Prejudice, and F (validation), and (d) higher scores on the MMPI Social Status and Intellectual Efficiency measures. When correlations were run between judging ability and these various tests, for the sexes separately, some male-female differences emerged, which are discussed in greater detail elsewhere (Cline, 1955a). Sex differences in interpersonal perception have been found in nearly all studies conducted by this investigator, indicating the necessity of separate analyses of data for male and female judges, and probably for "others" when they include both sexes.

C. The Generality of Judging Ability

Cline and Richards in 1957 began a new series of studies at the University of Utah in the area of interpersonal perception. The initial focus was on the problem of generality and its relation to a component type of analysis of global accuracy scores. A new set of filmed interviews in sound and color as well as much larger and varied group of judging instruments was developed. These instruments included

(1) *Behavior Postdiction* test. The development and format of this were similar to the same test cited in the previous study.

(2) *Sentence Completion* test. This is very similar to the previously discarded Multiple Choice Sentence Completion test described previously.

(3) *Adjective Check* (ACL). This evolved from the older previously mentioned Personality Word Card. In this case the judges were required to predict which of 20 adjectives the interviewee (or "other") had previously checked as being self descriptive. Only those adjectives which the interviewee had consistently checked on three different occasions as applying to himself were used as the "correct" items, and only those

adjectives never checked as self descriptive were used as the incorrect alternatives. Ten of the 20 adjectives were "correct" and the other ten "incorrect." In choosing the "correct" and "incorrect" adjective items for this judging test from the original 300 adjective pool, an attempt was made to choose socially desirable items as well as undesirable items in about equal numbers as correct and incorrect on the scoring key.

(4) *Opinion Prediction* test. Here the judge was required to predict how the interviewee (in the film) answered 20 MMPI items such as, " 'I fall in love easily,' True or False?" Only those MMPI items which the interviewee on three different occasions answered consistently the same way were used. It will be noted that each of the 25 interviewees had previously filled out answers to a selected group of 80 MMPI items on three occasions (several weeks apart). It was from this 80-item pool that the 20 consistently answered items were selected. An attempt was made to obtain ten of these items scored in the True direction and ten in the False direction and, additionally, that ten should have "favorable" type responses correct and ten "unfavorable" correct to prevent the judges from obtaining high accuracy scores merely by going through and check-ing only the "socially desirable" (or favorable) responses. Since this scale was constructed on the basis of each interviewee's responses, the particular items included in the scale differed from one interviewee to another.

(5) *Trait Rating* test. Here the judge was required to rate the interviewee on 50 traits (later reduced to 25) such as "cooperative," "impulsive," "efficient," "a leader," etc., on a six-point scale ranging from "very like" to "very unlike." The criterion was obtained from the mean of the ratings (on each of the items) of the interviewee by his friends and associates. Two other secondary types of criterion scores were also obtained: (1) the researcher's final summary ratings of the interviewee after evaluation of all of the test and interview data; and (2) the mean of the interviewee's own self ratings, done three times. This Trait Rating procedure was amenable to a Cronbach-type analysis into components.

Various revisions in scoring systems and format occurred with all of these instruments over time and a number of other instruments were developed, some of which are mentioned later in this paper.

While the problems in interpreting global scores of any kind were recognized, the different judging instruments were intercorrelated with results indicated in Tables V through XIII.[2] In addition, correlations

were obtained between single instruments and Total Judging scores. In this case, Total Judging scores were obtained by converting scores on each judging instrument to standard scores and adding across instruments. All correlations between individual instruments and Total Judg-

TABLE V

INTERCORRELATIONS AMONG JUDGING MEASURES FOR UNIVERSITY OF UTAH STUDENTS USING THE SIX- AND TEN-FILM FORMS

(N = 295)

	Ten-film form[a]			Six-film form[a]		
	Trait Rating	Adjective Check	Total Judging[b]	Trait Rating	Adjective Check	Total Judging[b]
Behavior Postdiction	.24	.04	.31	.36	.28	.39
Trait Rating		.31	.44		.41	.50
Adjective			.31			.44
		r .05 = .11			r .01 = .15	

[a] Note: Through procedures resembling item analysis, the number of films used during a judging test was eventually reduced to the "best six." And unless otherwise specified in the tables, it is this group of films which was used. Where the "10-film form," "20-film form," etc. are referred to it means that the particular research was performed before the final "best six" films had been selected.

[b] Part-whole correlations corrected for the contribution of the part.

TABLE VI

INTERCORRELATIONS BETWEEN ACCURACY SCORES ON DIFFERENT JUDGING MEASURES FOR UNIVERSITY OF UTAH STUDENTS

(N = 51)

	Trait Rating	Opinion Prediction	Adjective Check	Total Judging[a]
Behavior Postdiction	.16	.20	.37	.37
Trait Rating		.31	.13	.28
Opinion Prediction			.25	.32
Adjective Check				.36
r .05 = .27			r .01 = .35	

[a] Part-whole correlation corrected for contribution of the part.

TABLE VII

INTERCORRELATIONS AMONG JUDGING MEASURES FOR IDAHO STATE (MENTAL) HOSPITAL PATIENTS

(N = 32)

	Trait Rating	Adjective Check	Total Judging[a]
Behavior Postdiction	.41	.61	.67
Trait Rating		.31	.37
Adjective Check			.57
r .05 = .35		r .01 = .45	

[a] Part-whole correlation corrected for contribution of the part.

ing scores were part-whole correlations corrected for contribution of the part. In Table XV are given some of the reliabilities of the judging instruments. In Table XIV are listed the correlations between Trait

TABLE VIII

INTERCORRELATIONS AMONG JUDGING MEASURES FOR T GROUP MEMBERS

(N = 118)

	Trait Rating	Opinion Prediction	Adjective Check	Total Judging[a]
Behavior Postdiction	.28	.36	.36	.52
Trait Rating		.27	.03	.26
Opinion Prediction			.07	.47
Adjective Check				.35
r .05 = .17			r .01 = .23	

[a] Part-whole correlation corrected for contribution of the part.

TABLE IX

INTERCORRELATIONS AMONG JUDGING MEASURES FOR UNIVERSITY OF UTAH STUDENTS

(N = 50)

Measure	Trait Rating Total	Behavior Post-diction	Sentence Com-pletion	Opinion Pre-diction	Adjective Check	Total Judging[a]
Trait Rating Total	—					
Behavior Postdiction	.30	—				
Sentence Completion	.48	.47	—			
Opinion Prediction	.52	.50	.47	—		
Adjective Check	.65	.24	.58	.54	—	
Total Judging[a]	.63	.44	.63	.65	.66	—
r .05 = .27				r .01 = .35		

[a] Part-whole correlations corrected for the contribution of the part.

TABLE X

INTERCORRELATIONS AMONG JUDGING MEASURES FOR UTAH STATE (MENTAL) HOSPITAL STAFF

(N = 62)

	Behavior Postdiction	Trait Rating	Adjective Check	Total Judging[a]
Behavior Postdiction	—	.50	.22	.42
Trait Rating		—	.36	.48
Adjective Check			—	.31
Total Judging Score				—
r .05 = .25			r .01 = .33	

[a] Part-whole correlations corrected for the contribution of the part.

Rating Total score (where Trait Rating is a difference score judging instrument susceptible to a Cronbach-type component analysis, and the Trait Rating Total score is the global unanalyzed measure), its various components, and four other entirely different types of judging instru-

ments (Behavior Postdiction, Sentence Completion, Opinion Prediction, and Adjective Check—none of which were susceptible to a component analysis), plus a final Total Judging score, combining scores on all five instruments (through conversion to standard scores).

The conclusions drawn from these studies and data were that the evidence indicated that a general, global ability to judge others accurately could be meaningfully measured. However, the data in Table XIV suggested that this global ability involved a complicated dynamic process, or, more precisely, a factorially complex process that helped to clarify what was general in the global measure. Stereotype Accuracy appeared to

TABLE XI

INTERCORRELATIONS AMONG JUDGING MEASURES FOR UTAH STATE (MENTAL) HOSPITAL PATIENTS

($N = 56$)

	Behavior Postdiction	Trait Rating	Adjective Check	Total Judging[a]
Behavior Postdiction	—	.41	.32	.42
Trait Rating		—	.30	.44
Adjective Check			—	.31
Total Judging				—
$r\ .05 = .26$			$r\ .01 = .31$	

[a] Part-whole correlation corrected for the contribution of the part.

TABLE XII

INTERCORRELATIONS AMONG JUDGING MEASURES FOR UNIVERSITY OF UTAH STUDENTS USING 20-FILM FORM

($N = 23$)

	Behavior Postdiction	Trait Rating	Adjective Check
Behavior Postdiction	—	.36	.09
Trait Rating		—	.06
Adjective Check			—
$r\ .05 = .40$		$r\ .01 = .51$	

account for a large portion of this generality. However, after this component was eliminated, considerable generality remained, which appeared to be related mainly to the Differential Accuracy component. Thus these two components were to some degree independent. It appeared that one might be an accurate judge because one has an accurate stereotype or because one is able to predict specific differences between individuals or both. Bronfenbrenner et al. (1958) came to a nearly identical conclusion using quite different research procedures. They referred to the two components as "Sensitivity to the Generalized Other" and "Interpersonal Sensitivity" (or Stereotype Accuracy and Differential Accuracy in Cronbach's terminology).

It would appear paradoxical (or inappropriate) to conclude both that judging ability is general and that at the same time it consists of two relatively independent components. However it is probable that these seemingly conflicting conclusions mean that the ability to judge others

TABLE XIII

INTERCORRELATIONS AMONG JUDGING MEASURES FOR UNIVERSITY OF UTAH STUDENTS USING NINE-FILM FORM

$(N = 26)$

	Behavior Post-diction	Sentence Com-pletion	Opinion Pre-diction	Adjective Check I	Adjective Check II	Trait Rating I	Trait Rating II
Behavior Postdiction	—	.12	—.20	.04	.01	—.20	.06
Sentence Completion		—	.20	.13	.44	.32	.08
Opinion Prediction			—	.26	.42	.26	.16
Adjective Check I (no. of adjectives correctly predicted)				—	.41	.00	.19
Adjective Check II (% adjectives correctly predicted)					—	.29	.18
Trait Rating I[a] (sum of difference scores where criterion is 5 friends' ratings of "other"						—	.50
Trait Rating[a] (sum of difference scores where criterion is the "other's" self ratings)							—

$r\ .05 = .38$

$r\ .01 = .49$

[a] Signs of correlations have been changed so that they are consistent with other judging instruments.

accurately is factorially complex and that the independent components reflect this complexity. It is, therefore, still meaningful to talk about a general trait. There are numerous precedents for this in the history of psychology. For example, intelligence, which has been demonstrated to consist of several independent components, is still treated as a meaning-

ful general trait. In investigations of intelligence, even those studies indicating a *G* factor have typically also shown several independent group factors.

One possible criticism of the above research would relate to the

TABLE XIV
CORRELATIONS BETWEEN TRAIT RATING COMPONENTS AND OTHER JUDGING MEASURES
(*N* = 50)

Measures	Elevation	Differential Elevation	Stereotype Accuracy	Differential Accuracy	Trait Rating Total Score[a]
Elevation	—				
Differential Elevation	.12	—			
Stereotype Accuracy	.04	.01	—		
Differential Accuracy	.29	.29	.16	—	
Trait Rating Total Score[a]	.25	.24	.16	.25	—
Behavior Postdiction	—.05	—.23	.66	.01	.30
Sentence Completion	—.20	.08	.60	.25	.48
Opinion Prediction	.20	.00	.59	.28	.52
Adjective Check	.05	.23	.44	.56	.65
Total Judging (All Instrument)	.16	.10	.76	.53	.63
M	.1015	.1162	.7704	1.4392	2.4273
σ	.1154	.0589	.3911	.5759	.8139

r .05 = .27   *r* .01 = .35

[a] The correlations between the various components and the *Trait Rating* Total Score involve the total score with the contribution of the particular component under consideration eliminated. In computing the correlations between these component scores and the (*all instrument*) Total Judging score, however, it proved to be impossible to eliminate the contribution of the particular component under consideration, since the components represent the relative contribution of these components to the variance accuracy.

TABLE XV
RELIABILITIES OF JUDGING INSTRUMENTS

	5 Films vs. 5 films N = 50	10 Films vs. 10 films N = 20	6 Films only, N = 149[a]
Trait Rating	.72	.81	.60
Behavior Postdiction	.66		.37
Sentence Completion	.67		—
Opinion Prediction	.67		.59
Adjective Check	.79		.61
Total Judging	.71		.54

[a] Reliabilities for Behavior Postdiction, Opinion Prediction, and Adjective Check are computed by Kuder Richardson Formula 21. Trait Rating Reliability is a Spearman–Brown Coefficient based on the first three films—second three films split. Total Judging reliability was estimated using the formula presented by Mosier (1943).

possibility that the generality obtained was due to real similarity between persons judged. This view would suggest that the selection of the films on the basis of their discrimination of good judges from poor, as was done in this research, would result in a selection of films in which the "others" or interviewers were more similar to one another than they were to the "others" in the eliminated films.

In order to check this possibility, a special key was prepared for the Trait Rating, based on the mean rating on each trait of the criterion scores of the 10 films included in the study. Then an "Index of Similarity" was computed for each of the 10 best films included and for each of the 10 films eliminated on the basis of the earlier pilot study by scoring the *criterion* ratings for each film against this special key. The two groups of films were then compared with regard to this Index. The results of this comparison ($t = 1.694$, $df = 18$) fail to support a significant difference between the groups of films in this respect. The small "chance" difference that does occur, however, is in the expected direction. Generality of judging ability to the extent found in this study cannot, therefore, be explained on the basis of real similarity to one another of the persons judged.

D. Relationship between the Scoring System and Components

In Cline and Richard's early research it appeared that Cronbach's Stereotype Accuracy and Differential Accuracy components were going to provide a way out of the methodological morass in which most research had bogged down. However, additional work with Stereotype Accuracy and Differential Elevation demonstrated empirically that differences in the scoring system could spuriously increase, reduce, or even eliminate them, making Differential Accuracy difficult or impossible to interpret and generating misleading conclusions. Therefore a new measure termed "Interpersonal Accuracy" was developed (Richards and Cline, 1963) in the following manner.

Preliminary consideration of the components problem led to the development of a Belief-Values Inventory which required the judge to predict "other's" responses to 12 Likert-type items dealing with religious beliefs and values. The following is a sample item:

When in doubt, I have found it best to stop and ask God for guidance.

> A. Strongly agree
> B. Agree
> C. Neither agree nor disagree
> D. Disagree
> E. Strongly disagree

In the filmed interview the "other" had been asked direct questions about his attitudes toward religion.

Subsequent research with this instrument indicated that the scoring system had a differential effect upon the components involved in a Cronbach-type analysis. With items of the type included in the Belief-Values Inventory, there are two possible scoring methods. The first of these is to score "Strongly Agree" as 1, "Agree" as 2, etc., without regard to whether or not on that particular item, "Strongly Agree" is a pro-religious answer. The second possible scoring system is to score the most conventional proreligious response as 1, regardless of whether that answer is "Strongly Agree" or "Strongly Disagree." If the first of these scoring systems is used, the Stereotype Accuracy variance is large, but the Differential Elevation variance is made artificially small. On the other hand, if the second of these scoring systems is used, the Stereotype Accuracy variance is artificially reduced, while the Differential Elevation variance is maximized. These effects are illustrated by Table XVI, which presents the responses of three hypothetical persons to items of this type, with each item score presented in both of the two scoring systems. In Table XVI, the first hypothetical person always answered with the most conventional religious answer; the second hypothetical person always answered with the second most conventional religious answer, and the third hypothetical person always gave the middle or neutral response. In this table the consistency of responding by each person is, of course, somewhat exaggerated to make the point clear.

In addition to the effects of the scoring system on the variance of Stereotype Accuracy and Differential Elevation, several other things are apparent from Table XVI. If these three hypothetical persons were used as "others" in the films and the first scoring system were used (i.e., where "Strongly Agree" is always scored as 1, regardless of religious direction), no matter what degree of accuracy a judge attained in predicting their responses, the Differential Elevation correlation component could take no other value than .00. The Differential Elevation correlation term could have no relationship to other measures of judgment and might compel the erroneous conclusion that there is no generality of judging ability. Use of the second scoring system (i.e., where the most proreligious response is always scored 1 regardless of whether it is "Strongly Agree" or "Strongly Disagree") has a similar effect on the Stereotype Accuracy correlation and again could lead to an invalid conclusion that there is no generality of judging ability. It will also be seen from Table XVI that, if the first scoring system were used in a study of judging ability, Differential Elevation and also Elevation would reflect primarily the extent to which judges interpreted items in the same way as the "others," but

TABLE XVI

COMPARISONS OF RESPONSES OF THREE HYPOTHETICAL PERSONS TO BELIEF-VALUES-INVENTORY-TYPE ITEMS USING TWO DIFFERENT SCORING SYSTEMS[a]

Inventory item	Scores of three hypothetical people (A, B, C) when Strongly Agree is always scored 1			Stereotype Accuracy M
	A	B	C	
I believe in God.	1 (Strongly Agree)	2 (Agree)	3 (Neither	2
Religion is nonsense.	5 (Strongly Disagree)	4 (Disagree)	3 Agree	4
Prayers are answered.	1 (Strongly Agree)	2 (Agree)	3 nor	2
All churches should be closed.	5 (Strongly Disagree)	4 (Disagree)	3 Disagree)	4
Differential Elevation M	3	3	3	

Inventory item	Scores of three hypothetical people (A, B, C) when the most proreligious answer is always scored 1			Stereotype Accuracy M
	A	B	C	
I believe in God.	1 (Strongly Agree)	2 (Agree)	3 (Neither	2
Religion is nonsense.	1 (Strongly Disagree)	2 (Disagree)	3 Agree	2
Prayers are answered.	1 (Strongly Agree)	2 (Agree)	3 nor	2
All churches should be closed.	1 (Strongly Disagree)	2 (Disagree)	3 Disagree)	2
Differential Elevation M	1	2	3	

[a] Note: Person A always responds in the most conventional proreligious manner (i.e., "Strongly Agree," and "Disagree" where appropriate). Person C always chooses the neutral middle category, "Neither Agree nor Disagree."

that, if the second scoring system were used, Differential Elevation would be a measure of the judges' "sensitivity to individual differences" in over-all religiosity, and Elevation would be a measure of the judged average religiosity of the group of "others." Thus, the apparent paradox in Cronbach's formulation is resolved. All of this, taken together, strongly suggests that in investigations of accuracy of person perception, and particularly of its generality, neither of these scoring systems is by itself satisfactory, but rather that the first scoring system should be used in computing Stereotype Accuracy and its components and the second scoring system should be used in computing Differential Elevation and its components. It should be emphasized that, in the hypothetical example, the items differed greatly in their over-all degree of religiosity. In studies of accuracy of interpersonal perception, both of these would be important and yet either one or the other would inevitably be artificially eliminated if either scoring system alone were used.

These two scoring systems also have another effect which is not readily apparent in the hypothetical example since it eliminates all the persons-by-items interaction that would occur in a real problem. This effect is that the Differential Accuracy component and its correlation and variance constituents will all take on different values depending on which scoring system is used, and there is apparently no criterion which would indicate in a real problem which of these values are the most appropriate measures of judging ability. This indicated that under certain circumstances none of the values of Differential Accuracy and its constituents are particularly good measures of judging ability. Thus it was replaced with the new index of judging ability called Interpersonal Accuracy.

There is no difference score form of this measure; it consists only of a correlation term and a variance term. The correlation term is computed by determining the correlation between each judge's predicted values and the corresponding actual responses by "others" on individual items and then averaging across items (without converting these scores in terms of their discrepancy from item and person means as is the case with Differential Accuracy). Similarly, the Interpersonal Accuracy variance term involves the computation of the variance of each judge's predictions on individual items, averaged across items. In terms of Cronbach's scheme, it appears to be a linear combination of Differential Elevation and Differential Accuracy. It offers the strong advantage over other measures that it is invariant under changes of scoring system.

There are several additional considerations in the interpretation of the hypothetical example mentioned above. The first of these is that there is a strong general "religiosity" factor underlying the questic

In Cronbach's scheme, this general factor should be tapped by the Differential Elevation component, and this does occur under the scoring system in which the conventional religious answer is always scored 1. It might be objected, therefore, that the scoring system in which "Strongly Agree" is always scored 1 "randomizes" this general factor. This effect of this scoring system is, however, exactly the point of the hypothetical example; and it should be emphasized again that the scoring system in which "Strongly Agree" is always scored 1 is the only scoring system which permits the real differences in the average degree of endorsement of the items (Stereotype Accuracy) to appear. It should also be noted that the nature of the questions presented in the hypothetical example was intentionally made such as to emphasize the inappropriateness of this scoring system for the Differential Elevation component. It is a more usual procedure in studies of person perception to use rather heterogeneous groups of personality traits as items, and, as Cronbach points out, the assignment of scores to such items is frequently arbitrary. As a result, if such items are used, the scoring system will still have the effects outlined above, but in such complex and confusing patterns as to make the various component scores almost wholly uninterpretable (with the possible exception of the Total score). The fact that there is a strong general factor in the hypothetical example would contribute greatly to the clarity of interpretation of the Interpersonal Accuracy component, and the author is in agreement with Cronbach's position that several factorially pure sets of items analyzed separately are preferable to one factorially complex set of items treated in a global fashion.

If the argument is accepted to this point, it is still an open question whether the considerations outlined have any practical effect on investigations of accuracy of person perception. A study providing some information with regard to this point has been conducted. In this research 46 undergraduates, both male and female, at the University of Utah, predicted the responses of six standard "others," presented through the filmed interview procedure, on the Belief-Values Inventory. Details of the experimental procedure of using these filmed interviews are presented elsewhere (Cline and Richards, 1960). The predictions of these judges were scored twice against the criterion, once with each of the two scoring systems discussed above, and the various judgment scores intercorrelated. When this program is used all correlation terms are expressed in terms of Fisher's z. Results of both of these analyses are presented in Table XVII. In this table, correlations above the diagonal were obtained when "Strongly Agree" was always scored 1 regardless of whether or not it represented a proreligious answer, and correlations below the diagonal were obtained when the most conventional proreligious answer was

TABLE XVII

INTERCORRELATIONS OF BELIEF-VALUES INVENTORY COMPONENTS FOR EACH OF TWO SCORING SYSTEMS[a]

	Total	Elevation	Differential Elevation	Stereotype Accuracy	Differential Elevation z[b]	Stereotype Accuracy z	Interpersonal Accuracy z	Differential Elevation σ²	Stereotype Accuracy σ²	Interpersonal Accuracy σ²
Total[c]		.00	.24	.37	.21	.72	.84	−.40	.46	.07
Elevation	.55		.03	−.11	−.06	−.07	.11	.03	−.16	.18
Differential Elevation	.38	.36		.13	.80	.06	.30	−.41	.08	−.20
Stereotype Accuracy	.01	.11	−.17		.06	.95	.56	−.15	.58	.35
Differential Elevation z	.67	.46	.80	.01		.05	.25	.06	.08	−.03
Stereotype Accuracy z	.19	.02	.05	.53	.18		.51	.10	.45	.32
Interpersonal Accuracy z	.84	.62	.72	.08	.86	.23		−.16	.37	.33
Differential Elevation σ²	.28	.42	−.07	.21	.33	.26	.25		.06	.47
Stereotype Accuracy σ²	−.11	−.17	.22	−.71	.10	.17	.05	−.10		.40
Interpersonal Accuracy σ²	.07	.33	−.21	.20	.22	.22	.33	.96	−.13	

a Note: Correlations above the diagonal were obtained when Strongly Agree was always scored 1, and correlations below the diagonal were obtained when the most conventional religious answer was always scored 1. r .05 = .29; r .01 = .37.

b All z components are Fisher's z transformations of Pearson correlations.

c All correlations between Total and difference score components are corrected part-whole correlations.

always scored 1. On the basis of either of these two groups of correlations alone, one would have to conclude that there is no consistent pattern of generality in judging ability, particularly as reflected in the three correlation measures, but rather that there appear to be two relatively independent factors measured respectively by the Differential Elevation correlation term and the Stereotype Accuracy correlation term, thus confirming the previous results of Bronfenbrenner and his associates (1958) and Cline and Richards (1960).

A further analysis of these data was made, however, in which judgment scores were intercorrelated across scoring systems in such a way that each component was scored most appropriately. More specifically, Stereotype Accuracy and its correlation and variance terms were computed using the scoring system where "Strongly Agree" was always scored 1, and Differential Elevation and its components and Elevation were computed using the scoring system where the most proreligious answer was always scored 1. Results are presented in Table XVIII. In this table, there is a consistent pattern of a significant degree of generality across the correlation terms of all components, thus suggesting that judging ability is, to some degree, a general trait.

These results indicate that the scoring system may be an important artifact in investigations of the generality question when components of accuracy scores are used, and therefore strongly support the argument advanced in the hypothetical example discussed earlier, though it is still most important to avoid overgeneralization from these results.

E. Effect of Varied Stimulus Information on Judging-Accuracy Scores

The next focus was on the fate of the various components when the amount of information about the stimulus person or "other" is systematically increased.

Using the Belief-Values Inventory (judging instrument) and the standard set of six filmed interviews, one group of 95 judges made their judgments on the basis of seeing and hearing the filmed interviews and were termed the full information group. A second group consisted of 50 persons who filled out the Belief-Values Inventory without seeing the films at all, only knowing the age, sex, marital status, and number of children of the six stimulus persons. The third group consisted of 58 judges who filled out the same inventory only twice, once as they thought it would apply to a typical American adult male and once for a typical adult American female. Thus, in these last two groups the judges were required to make judgments solely on the basis of stereotypes, though these stereotypes differed in the amount of information on which they

TABLE XVIII

INTERCORRELATIONS OF BELIEF-VALUES INVENTORY COMPONENTS WHEN EACH COMPONENT IS SCORED APPROPRIATELY[a]

	Total	Elevation	Differential Elevation	Stereotype Accuracy	Differential Elevation z	Stereotype Accuracy z	Interpersonal Accuracy z	Differential Elevation σ^2	Stereotype Accuracy σ^2	Interpersonal Accuracy σ^2
Total[b]										
Elevation	.55									
Differential Elevation	.38	.36								
Stereotype Accuracy	.37	.93	.25							
Differential Elevation z	.67	.46	.80	.40						
Stereotype Accuracy z	.72	.82	.22	.95	.38					
Interpersonal Accuracy z	.84	.62	.72	.56	.86	.51				
Differential Elevation σ^2	.28	.42	−.07	.48	.33	.43	.25			
Stereotype Accuracy σ^2	.46	.70	.18	.58	.26	.45	.37	.48		
Interpersonal Accuracy σ^2	.07	.33	−.21	.35	.22	.32	.33	.96	.40	

[a] Note: All z components are Fisher's z transformations of Pearson correlations. r .05 = .29; r .01 = .37.

[b] All correlations between Total and difference score components are corrected part-whole correlations.

were based. The results (see Richards, Cline, and Rardin, 1962) for the Interpersonal Accuracy z and Total Error score indicated that the judges in the "full information" group were significantly superior both at rank ordering the stimulus persons in terms of conventional religious values *and* at predicting exactly the responses of these same stimulus persons. Results for Interpersonal Accuracy variance indicated that differentiation among stimulus persons increased as the amount of information provided about them increased. Contrary to what would be expected, Stereotype Accuracy z varied significantly among the three groups while Stereotype Accuracy variance decreased as the amount of information increased. This probably indicates that Stereotype Accuracy is a measure of considerable complexity and contrary to previous opinion probably involves some degree of sensitivity to individual persons. Additional research is needed on this point.

F. Components of Judging-Accuracy Scores Applied to the Clinical *versus* Statistical Prediction Controversy

A controversy relevant to some of the basic issues in interpersonal perception relates to clinical *versus* statistical prediction. Meehl (1955) in surveying all studies extant in which the relative predictive accuracies of clinical and actuarial techniques were compared concluded that no published study indicated a clear superiority for clinical procedures. More recently, Holt (1958) has suggested that these conclusions resulted from the fact that the studies on which they were based involved kinds of predictions to which actuarial procedures are better suited than are clinical procedures. Richards (1963) has suggested a conceptual scheme to clarify this issue using a modification (Cline and Richards, 1960) of the analytic model for person perception scores developed by Cronbach (1955).

Cline and Richards (1962) conducted an empirical study to compare the clinician and the actuary, using a component analysis of the judging accuracy scores. They argued that in investigations of predictive accuracy using the component method of analysis, Interpersonal Accuracy, and particularly its correlation subscores, is the most appropriate measure of sensitivity to individual differences. Cronbach states that Stereotype Accuracy is not a measure of this type of sensitivity, but rather of the accuracy of prediction of the norm for the group of persons about whom the predictions are made (and, in some circumstances, for people in general). It may be argued that while all components are important, the central goal of "real life" clinical prediction is to achieve accuracy on components measuring sensitivity to individual differences. On the other hand, as Richards (1963) points out, the accuracy of actuarial prediction

can only be due to the fact that actuarial tables have an accurate stereotype. Since this is the case, studies which indicate that actuarial prediction is superior to clinical prediction must indicate that actuarial tables have a more accurate stereotype than clinicians do; in other words, a higher score on the Stereotype Accuracy correlation term. This has been demonstrated in research (Halbrower, 1955) cited by Meehl (1956) to support his position.

On those components measuring "sensitivity to individual differences," however, actuarial prediction is at an inherent disadvantage in that it must predict all members of a given class of persons to have obtained the same score. On those components, therefore, any accuracy achieved by actuarial procedures must be due to differences between classes. The clinician has no such restriction on his predictions: he can easily predict differences within classes, thus increasing the range and thereby his possible accuracy on components measuring sensitivity to individual differences. It is obvious that if he does predict differences within classes, it may either increase or decrease his accuracy. However, the author knows of no reason to suppose that it would always result in a decrease, and concludes that in any given case it is an empirical question whether accuracy is increased or decreased.

Much of the above was, of course, recognized explicitly by Meehl in his original book, but he could find no evidence that the clinician's greater freedom resulted in greater accuracy. Probably two factors account for this. The first of these is that no studies of clinical *versus* statistical prediction have made either an explicit or implicit comparison of the various components of accuracy scores. This is not a criticism, since no such analytic scheme was available at the time most of these studies were conducted. The second major reason for Meehl's findings was, in the opinion of Cline and Richards, that many, or most, studies comparing clinical and statistical prediction have used measures which reflect mainly Stereotype Accuracy, which would favor the actuarial procedures. This is particularly clear in the case of the "Q correlation," which is commonly used in studies of clinical prediction. Q correlation is a within-persons correlation and is therefore based on differences between items. Differences between items, however, are defined as Stereotype Accuracy both in Cronbach's model and in the author's modified model.

On the basis of the rationale outlined above, Cline and Richards developed two hypotheses for the application of this scheme of analysis into component scores to the clinical *versus* statistical prediction studies. These hypotheses were that (1) in all studies using an acceptable actuarial prediction procedure, actuarial prediction will be superior to clinical on the Stereotype Accuracy correlation component; and (2) in some such

studies clinical prediction will be superior to actuarial on the Interpersonal Accuracy correlation subscore. The purpose of the present study was to obtain data illustrating the use of component scores in a study of clinical *versus* statistical prediction, and to demonstrate that it is possible to obtain results favoring actuarial prediction on many of the commonly used measures of accuracy, but favoring clinical prediction on Interpersonal Accuracy.

It is true, of course, that the accuracy of actuarial prediction on Interpersonal Accuracy can always be increased by more and more accurate division into classes of persons. This, however, does not invalidate the basic argument, since it would still be an open question whether or not the accuracy of actuarial prediction on this component could be increased with reasonable economy to a level equal to or higher than the accuracy of clinical prediction. Most actuarial prediction procedures have used relatively crude classifications, achieved with considerable effort.

In this study the Ss were 56 students, male and female, from undergraduate psychology classes at the University of Utah. While it is hazardous to generalize from such a population of naïve untrained students to a population of highly trained clinical psychologists, this is a conservative error, likely to penalize clinical prediction. If this is the case, then any superiority over actuarial prediction found for these judges can be generalized to trained clinicians.

In this research, the "clinical predictions" were made after the Ss had seen sound color movies of brief interviews of six different persons. These films were selected from a large group of 25 films on the basis of procedures resembling item analysis which are reported in detail elsewhere (Cline and Richards, 1960). The six films used presented the interviews of three males and three females, and the persons interviewed were selected so that they were diverse in age and background. With particular regard to the present research, they were selected so that they were heterogeneous in their self-reported acceptance of conventional religious beliefs and practices. The filmed interviews included direct questions about views on religion.

The task for Ss in this experiment was to fill out one judging instrument, the Belief-Values Inventory, as they thought the persons who were interviewed in the films had filled it out. The Belief-Values Inventory consisted, as previously mentioned, of Likert-type items dealing with personal values in the religion area. There were 12 items for each of the six films, and the same 12 were used for every film.

Six scores for each S were computed on the basis of this instrument. Five of these were Total Error Score, Stereotype Accuracy z, Stereotype Accuracy variance, Interpersonal Accuracy z, and Interpersonal Accuracy

variance. In the present study, the Interpersonal Accuracy z score measured primarily the success with which these six persons were rank ordered in terms of their over-all "religiosity."

The sixth score calculated on the basis of the Belief-Values Inventory is called Q-Z, still another expression of correlation in terms of Fisher's z. In determining the value of this score for a given judge, the responses to the items predicted by the judge for each of the six persons in the films were correlated with corresponding actual responses for the six persons individually. In other words, it is a "within persons" correlation, highly similar to correlations based on Q sorts. These correlations were then converted to Fisher's z and averaged across the six persons. This score was used in this study because it corresponds closely to a common measure of accuracy of clinical prediction, and therefore makes possible comparisons between the present research and earlier studies of the clinical and statistical prediction problem.

At the start of each judging session, the Ss filled out forms, dealing with necessary background information, after which the film of the first interview was shown. The projector was stopped, and the Ss made their predictions or judgments about the first film. As soon as all Ss completed their judgments, the film of the second interview was shown. This procedure was repeated until all six films had been shown and all judgments completed.

In previous studies comparing clinical and statistical prediction a variety of methods had been used. What these methods appear to have in common is the establishment of norms for some group or set of subgroups and then the derivation of predictions about members of an independent group. Meehl (1955) has suggested a further requirement, that both prediction procedures use exactly the same information, but Estes (1961) has pointed out that none of the studies cited by Meehl actually met this requirement. This "defect" did not prevent Meehl from deriving important and legitimate conclusions, and therefore should not prevent other investigators from deriving important and legitimate conclusions from similar studies.

This implies that the most appropriate procedure to use for statistical prediction in this study would be to establish norms for the Values-Belief Inventory on the basis of an independent group and use these norms in predicting the responses of the smaller set of persons appearing in these filmed interviews. Accordingly, the following procedure was used. The group used in defining the norms consisted of 56 persons who were interviewed and filmed at two different periods during the course of the author's on-going research project on interpersonal perception. The 56 films did not, of course, include the six used in the present

research. On the basis of a "total religiosity" score on the Belief-Values Inventory, these films were divided into two groups on the basis of a median split, one "high" on religiosity (i.e., above the median) and one "low." The "total religiosity" score is defined as the sum of the scores on the 12 Belief-Values Inventory items when the scoring is adjusted so that a high score on each item represents acceptance of conventional religion. Norms were then computed for each of these two groups by averaging the responses of persons falling into each group to the 12 Belief-Values Inventory items.

The next step was to compute the "total religiosity" score for the six films used in the present research. The responses of the six persons appearing in the films were then "predicted" from the norm groups by determining if their individual "total religiosity" score would have placed them above or below the median of the norm group, and predicting that they had responded with the appropriate 12 average responses. Two of the persons used in this research fell above the median, and four fell below the median. It would appear that the information used by the actuarial prediction procedure is equivalent to that used by the clinical prediction procedure, although it is not identical.

There is one important thing that should be noted about this procedure: this is, if a group of similar studies were conducted, the scores for statistical prediction would be subject to sampling variation. Unfortunately, it is impossible to estimate what this variation would be and use it in statistical tests of significance of difference between clinical prediction and statistical prediction. This makes it necessary that the statistical test used be the t test for significance of deviation from a point of value (McNemar, 1955), with the error estimate involving only variation in clinical predictions (i.e., variation among judges). It should be emphasized that this is not a conservative error, since it would result in rejecting a true null hypothesis too often. Accordingly, it was decided to require the .01 level of significance before rejecting the null hypothesis rather than the conventional .05 level. All t tests reported in this study are two-tail tests.

The comparisons between clinical prediction and the actuarial prediction procedure on the six scores used in this study are presented in Table XIX. The results of these comparisons indicate that clinical prediction was significantly poorer on the Total Error Score, the $Q-Z$ score, and the Stereotype Accuracy correlation term, but was significantly superior to statistical prediction on the Interpersonal Accuracy correlation score. The results of these comparisons can be summarized as indicating that the statistical prediction procedure had a more accurate "stereotype" than did the Ss who made the "clinical" judgments, but

that the "clinician" Ss were more successful than the statistical prediction procedure in judging differences among persons. The comparison of the procedures on the variance components as expected indicates that the clinicians differentiated much more among persons than did the actuarial procedure. These results, taken together, reveal a complex pattern not reflected in the global Total Error Score, and thus strongly support the usefulness of component scores analysis in studies of clinical and statistical prediction. This suggests that the present research may be generalized to other such studies.

TABLE XIX

COMPARISON OF CLINICAL AND STATISTICAL PREDICTION ON COMPONENTS OF ACCURACY SCORES[a]

Score	Clinical prediction mean ($N = 56$)	Standard error of mean	Statistical prediction score	t	p
Total error score	1.22	.0334	.61	18.26	.001
Stereotype Accuracy Z	1.09	.0494	2.09	20.24	.001
Stereotype Accuracy variance	.45	.0267	.39	2.25	NS
Interpersonal Accuracy Z	.89	.0227	.73	7.05	.001
Interpersonal Accuracy variance	1.20	.0601	.45	12.48	.001
Q-Z	.82	.0251	.89	2.79	.01

[a] Note: Owing to the fact that no estimate of the sampling variation in statistical prediction is available, the statistical tests would reject a true null hypothesis too often. Therefore, a $p > .01$ rather than $> .05$ in the significance tests is required.

One purpose of this research was to demonstrate that it is possible to obtain results favoring actuarial prediction on many of the commonly used measures of accuracy, but also favoring clinical prediction on Interpersonal Accuracy, and this goal has clearly been achieved. While the results of the present study should not be overgeneralized, they should not be undergeneralized either. In particular, this statistical prediction procedure, while limited, is comparable to those used in many, or most, previous studies of clinical and statistical prediction.

In addition, the present research clearly suggests that, contrary to common opinion among clinicians, the activity at which clinicians are most likely to exceed the actuary is not making predictions about a "unique individual." This does not mean that there are no dramatic

instances in which a clinician makes successful predictions about a single individual by deviating markedly from actuarial expectancy, but rather that if the average clinician's major activity is making predictions within individual persons, his average accuracy will at best be equal to and at worst considerably less than the accuracy of statistical prediction since his stereotype tends to be less accurate. On the other hand, the present research does suggest that the most appropriate activity for clinicians is predicting differences among persons who are grouped into the same class by statistical prediction, and it is easy to suggest many ordinary activities of clinicians that are of this type. An example of this might be a mental hospital in which 10 patients had roughly the same statistically derived "predicted benefit from psychotherapy" scores, but which had such a limited staff that only three of these patients could actually be given psychotherapy. The results of this study suggest that the clinician might be quite successful in picking the three "best bets" although all 10 had been given an "equal" rating by the statistical prediction procedure.

G. Comparisons of Groups versus Individuals in Making Judgments about Others

A question with a somewhat different direction—whether a group of people can collectively make more accurate judgments than a single individual—was asked. The rationale of this experiment (Cline and Richards, 1961) grew out of the recent survey of studies comparing group performance and individual performance made by Lorge, Fox, Davitz, and Brenner (1958). The general conclusion of this survey was that a group, on almost any task, will perform better than a typical individual, but not necessarily better than a superior individual on the task in question. This finding is true whether the "group performance" is made by a genuine group or is merely a statistical combination of several independent individual performances. An unresolved question is the degree to which these findings can be attributed to a reduction in the variability of the group performance.

The trend of the studies cited in this survey suggested the hypothesis to be tested in this experiment. This hypothesis was the following:

The accuracy of predictions (about the behavior of other persons) made by a group of persons arriving at a consensus prediction through group discussion will be significantly greater than the average accuracy of the predictions made by the individuals composing the group. The average accuracy of the prediction made by the individuals composing the group will also be significantly less than the accuracy of an "artificial group" (composed of pooled independent judgments for each item) and

also less than the accuracy of prediction of the best individual among the individuals composing the group.

A secondary question relates to the presence or absence of a consistent pattern of superiority in accuracy among predictions made by the best individual judges, consensus groups, and artificial groups composed of pooled independent judgments for each item.

The subjects were 186 students, both male and female, in the introductory psychology classes at the University of Utah. The procedure involved the presentation of six filmed interviews or "standard others."

In this study, two prediction instruments were used. The first of these was the Adjective Check (ACL), which required the subject to determine which of a pair of adjectives the interviewee had checked as being descriptive of himself. There were 20 such pairs for each of the six films, making a total of 120. The score on the ACL was the number correct. Thus the ACL is similar to a forced-choice rating procedure. The second instrument used was the Belief-Values Inventory (BVI), already discussed.

There were 12 BVI items for each film or interview. Several different scores were computed from judges' responses to this instrument using a program developed for the IBM 650 computer. The first of these was a total score, which was based on the average of the squared discrepancies (using the one-to-five point scale) between predicted responses by each judge for each interviewee, and actual responses of each interviewee. This is an error score and in order to make these scores comparable with other scores used in this study, the scores were converted to accuracy scores through a standard score transformation, setting the mean equal to 50 and standard deviation equal to 10.

The second two BVI scores are components of what Cronbach (1955) has called "Stereotype Accuracy": (a) correlation between each judge's predicted item means and obtained item means, converted to a Fisher's z, and (b) the variance of each judge's predicted means. Cronbach has demonstrated these two scores to be the two parameters in Stereotype Accuracy when the criterion is held constant, and they permit independent evaluation of the effect of grouping on accuracy and on variability of prediction in this study. The last two scores on the BVI are measures of Interpersonal Accuracy.

The 186 subjects in this experiment were divided into 62 three-person groups. The division was made at the time the experiment was conducted, and most groups consisted of three persons seated next to each other in the experimental room. Group composition in terms of sex of group members was roughly random. The Ss saw each film and then completed the judging instruments independently. Next, they joined

together in a group discussion fashion and proceeded to arrive at a consensus judgment for the items on the judging instruments without referring back to their earlier independent judgments.

The "artificial group" judgment (pooled independent judgments for each item) was derived from the individual judgments of the group members. Thus, on the ACL, the artificial group judgment was determined on the basis of a "majority vote" of the judges on each item (by inspecting their individual judging protocols). On the BVI, it was calculated by determining the average of the values predicted by the three judges for each interviewee on each item. It is important to emphasize that this artificial group is only a statistical combination of the original independent judgments for each item.

The average accuracy of individuals composing the group was obtained by computing the mean of the total accuracy scores of the three individuals who made up each group. This is not the same as the artificial group procedure where the actual item-by-item predictions of the three group members were averaged rather than their total accuracy scores.

The best judge in each group was selected on the basis of his accuracy scores. In interpreting the results of this study, therefore, it is important to note that this selection was done on an after-the-fact basis, thus maximizing accuracy scores for this condition by capitalizing on chance. It would, therefore, be impossible for a best judge selected in advance to obtain a higher score than this, and such a best judge would, in fact, probably score somewhat lower, since some error would be involved in any advance selection. The best judges were selected independently for the ACL and the BVI and therefore were not necessarily the same person on the two different instruments. On the BVI, however, the best judges, selected on the basis of total score, were also used as best judges in making the comparisons involving the other scores derived from this instrument.

The mean and standard deviations for each judgment procedure on each judgment score are presented in Table XX. In Table XX, all scores are accuracy scores. Since total score on BVI is based on error score, in Table XX this judgment score is transformed to a standard score distribution with mean = 50 and standard deviation = 10.

As a first step in the statistical analysis of these data, over-all F tests were calculated for each of the judgment scores separately. The results of this analysis are presented in Table XXI. No test for homogeneity of variance was made before calculation of these F tests. This procedure was followed because the recent work of Boneau (1960) strongly suggests that F is not significantly affected by heterogeneity of variance if the sample

TABLE XX
MEANS AND STANDARD DEVIATIONS OF JUDGMENT SCORES

	Average of individuals composing the group	Best judge	Three-person group consensus	Artificial group derived by pooling three independent judgments
Adjective Check List				
$\overline{X}$	97.27	101.66	102.52	103.32
σ	3.51	3.91	3.95	4.55
Belief-Values Inventory Total				
$\overline{X}$	43.29	53.92	49.47	52.87
σ	8.61	8.31	11.16	8.02
Stereotype Accuracy z				
$\overline{X}$	1.19	1.44	1.28	1.41
σ	.28	.37	.45	.37
Stereotype Accuracy variance				
$\overline{X}$	.35	.40	.31	.30
σ	.14	.22	.15	.13
Interpersonal Accuracy z				
$\overline{X}$	.90	1.01	1.00	.98
σ	.12	.14	.16	.15
Interpersonal Accuracy variance				
$\overline{X}$	1.09	1.06	1.06	.91
σ	.23	.28	.31	.25

TABLE XXI
RESULTS OF OVER-ALL F TESTS FOR JUDGMENT SCORES

Judgment score	Between-variance $(df = 3)$	Within-variance $(df = 244)$	F
Adjective Check List Total	451.82	13.82	32.62[a]
Belief-Values Inventory			
Inventory total	1423.02	84.33	16.87[a]
Stereotype Accuracy z	.8633	.1432	6.03[a]
Stereotype Accuracy variance	.1333	.0282	4.72[b]
Interpersonal Accuracy z	.1633	.0213	7.67[a]
Interpersonal Accuracy variance	.3900	.0754	5.17[b]

[a] $p > .01$.
[b] $p > .05$.

TABLE XXII

TESTS FOR SIGNIFICANCE OF DIFFERENCE BETWEEN INDIVIDUAL MEANS FOR EACH JUDGMENT SCORE[a]

Judgment score	Average of individual vs. best judge	Average of individual vs. group consensus	Average of individual vs. artificial group	Best judge vs. group consensus	Best judge vs. artificial group	Group consensus vs. artificial group
Adjective Check List						
Total	4.39b	5.25b	6.05b	.86	1.66c	.80
Belief-Values Inventory						
Total	10.63b	6.18b	9.58b	4.45b	1.05	3.40c
Stereotype Accuracy z	.25b	.09	.22b	.16c	.03	.13c
Stereotype Accuracy variance	.05	.04	.05	.09b	.10b	.01
Interpersonal Accuracy z	.11b	.10b	.08b	.01	.03	.02
Interpersonal Accuracy variance	.03	.03	.18b	.00	.15b	.15b

[a] Note: Entries in this table represent the absolute difference between groups without regard to the direction of the difference.
[b] $p < .01$.
[c] $p < .05$.

sizes are identical and relatively large, i.e., 20. Both of these conditions hold in the present study. It is also known that available tests for homogeneity of variance are affected too much by other variables than that involved in the null hypothesis to justify their use prior to an analysis of variance (Box, 1953).

Since all of the F tests in Table XXI are significant at or beyond the .01 level of confidence, a test for significance of difference between individual means was made. This test was made using the multiple range test (Li, 1957, p. 238), which is an appropriate procedure for making "post-mortem" type comparisons between individual means after an over-all F test has been made. Briefly, the multiple range test involves computing a value which represents how large the difference between two means must be in order to be significant at a stated level, and then comparing the obtained difference to this value. Results of this analysis are summarized in Table XXII.

On each of the four accuracy measures, the best judge and both group judgments are significantly superior to the average of the individuals composing the group. Thus the major hypothesis of this experiment was confirmed. There is no consistent pattern of significant differences among the first three procedures mentioned above. As would be expected, on the two scores representing the amount of variability in predictions, the artificial group mean tends to be lower than the means of the other three procedures. This tendency is significant, however, only for the Interpersonal Accuracy variance score. It is somewhat surprising to find that the artificial group (or pooled independent judgments of items) is superior to the best judge on the ACL. The interpretation of this finding seems to be that if both other judges disagree with the best judge, they are more likely to be right than is the best judge. If, on the other hand, only one of the other judges disagrees with the best judge, he is more likely to be wrong than is the best judge.

This study clearly implies that satisfactory ratings are least likely to be obtained from a single unselected individual. In exploring further implications of these results for an operational rating setup, several other considerations enter. The first of these is that typically the best judge would be difficult to select on an *a priori* basis, and (because of selection error) best judges selected *a priori* would probably score lower than the best judges used in this study. Since each of the group procedures produces results roughly equivalent to those for the best judge selected on an after-the-fact basis, an extensive (and expensive) effort to identify best judges and use them as raters would appear to be unnecessary.

The second consideration involved in applying these results is that by far the most time in this experiment was consumed in arriving at

consensus judgments through group discussion, a finding which one would certainly expect to generalize to other situations. Since the artificial group (or pooled independent judgments of items) procedure produced results as good as or better than the results produced by the consensus judgment, and required much less time, it would appear to be most appropriate when accuracy and time are both considered. Thus, the best procedure for using ratings in many applied situations would be to obtain several independent ratings from different raters for each ratee, and then combine these ratings statistically into a single rating. It should be noted, however, that the superiority of the artificial group in terms of time required (and therefore expense) might disappear if only a single summary rating were required rather than the many relatively specific judgments required by the experimental procedure used in this study.

A limitation to these conclusions is the fact that each rater in this experiment was basing his ratings on the same or identical information (i.e., seeing the same movies of the interviews). If different raters are basing their ratings on different information, some other procedure involving the sharing of this information might be superior.

In addition to the practical implications outlined above, these results present at least two more additions to previous psychological research. The first of these is the demonstration through both the Stereotype Accuracy correlation term and the Interpersonal Accuracy correlation term of the BVI that accuracy is increased through grouping independent of a reduction in variability (see Table XX). Unlike the other results of this experiment, this would not necessarily be expected on the basis of previous studies comparing group and individual performance, although it certainly is consistent with previous studies. The second major addition is related to the current controversy in the interpersonal perception literature over the relative merits of various different types of accuracy scores (Cronbach, 1955). In the current study the total score on the ACL, the total score on the BVI, and the Stereotype Accuracy and Interpersonal Accuracy correlation terms all gave consistent results and, more important, results which make sense in terms of previous research comparing group and individual performance.

H. A Factor Analytic Study of the Interpersonal Accuracy Component

The Interpersonal Accuracy judging score component has been discussed previously and used in a variety of settings. As already mentioned the correlation term of this component is regarded as most closely related to what is ordinarily meant by accuracy of interpersonal perception. This term is computed by determining the correlation between

predicted, or judged, scores on individual items, and then averaging across items after transforming the item correlations to Fisher's z values. Where a strong general factor underlies the items, such as religiosity, this component measures the extent to which the persons judged are rank ordered accurately on that factor. Review of the results of Cline and Richard's various studies of this component has proven them complex and at times confusing. While it has responded at the *group* differences level to experimental manipulation in meaningful ways, it has tended to show little generality at the *individual* differences level across different measures (i.e., different sets of items or persons judged).

One possible interpretation of this pattern of results is that this component has validity but not reliability. Such a paradoxical result would be possible if the component were factorially complex, and this possibility suggested the hypothesis for the present study. This hypothesis is that when the individual item z's are treated as scores, intercorrelated, and rotated to an analytic solution, there will be several group factors rather than a strong general factor.

In order to determine whether this might be the case, the following study was undertaken in which 129 University undergraduate students participated as Ss. They were shown the standard six filmed interviews and given the judging tests (for each film), each of which was susceptible to an analysis in terms of components. The titles of the judging instruments and the characteristics underlying their item content are given below:

(1) Personal Practice Questionnaire (participation in religious activities),

(2) Activities Preference Inventory (femininity of interests),

(3) Interpersonal Relations Inventory (socialization).

All of the items from these scales were put into a Likert format. On the Activities Preference Inventory and the Interpersonal Relations Inventory, five responses were provided for each item, ranging from "Strongly Agree" to "Strongly Disagree." Scores from 1 to 5 were assigned to these responses so that "Strongly Disagree" always received a score of 5. On the Personal Practices Questionnaire four responses were provided for each item varying from "Frequently" to "Never." Scores from 1 to 4 were assigned to these responses so that "Never" always received a score of 4. There were ten Likert items per measure.

For each S, his scores on a given instrument were his 10 individual item Interpersonal Accuracy correlation values. In other words, on each item his score was the Z conversion of the correlation between his predicted scores for the six stimulus persons on that item and their cor-

responding "true" scores. It was thus a measure of his success in rank ordering the stimulus persons on that item.

These scores were intercorrelated and factor analyzed for each instrument separately using a principal components solution and varimax orthogonal rotation. Unity was placed in the diagonal and all components

TABLE XXIII

ROTATED LOADINGS FOR FACTOR ANALYSIS OF INDIVIDUAL ITEM ACCURACY SCORES ON THE PERSONAL PRACTICES QUESTIONNAIRE[a]

Item	Factors			
	A	B	C	h^2
1. Within the past few years I have attended religious services.	.80	—.26	—.04	.71
2. I have taught "Sunday School" or given another religious class or group within the last two years.	.00	.53	.18	.31
3. Within the last two years I have invited someone not of my faith to attend religious services with me.	.69	—.05	.18	.51
4. Within the last few years I have discussed religious topics with my friends.	.34	—.33	.21	.27
5. Within the past two years I have prayed.	.22	.79	—.04	.67
6. In my family we have the practice of having family prayer.	.53	.05	.32	.39
7. Within the past two years I have spent periods of time in private religious thought and meditation.	—.25	.65	—.10	.50
8. Within the past two years I have read the Holy Scriptures.	.15	—.17	.76	.63
9. Within the past two years I have read books, magazine articles, etc. with religious themes or "inspirational messages."	.04	.23	.79	.68
10. Within the past two years I have given money (or donations) to a church or religious group.	.64	.20	—.06	.45

[a] Note. For each item four choices were provided ranging from "Frequently" to "Never."

with an eigenvalue greater than 1.00 extracted. Kaiser (1960) presents in detail the rationale for this factoring procedure.

The rotated factor loadings and communalities for the Personal Practices Questionnaire are presented in Table XXIII, for the Activities Preference Inventory in Table XXIV, and for the Interpersonal Relations Inventory in Table XXV.

The implications of these results are clear. Interpersonal Accuracy

is a complex measure, involving several orthogonal factors rather than a large general factor and the factorial complexity is a much more important finding than the details of the individual factors. This indicates that accuracy probably should not be interpreted as a single unidimensional trait, but rather as a group of relatively independent traits. It should be emphasized that even though this is true one can still talk

TABLE XXIV

ROTATED LOADINGS FOR FACTOR ANALYSIS OF INDIVIDUAL ITEM ACCURACY SCORES ON THE ACTIVITIES PREFERENCE INVENTORY[a]

Item	Factors				
	A	B	C	D	h^2
1. I would like the work of a clerk in a large department store.	.54	—.37	—.23	—.16	.51
2. I like adventure stories better than romantic stories.	—.11	.58	—.24	.29	.49
3. I would like to be a soldier or a WAC.	—.09	.08	—.56	.05	.33
4. I very much like hunting.	—.02	.03	—.19	—.76	.62
5. I think I would like the work of a librarian.	—.04	—.03	—.22	.67	.50
6. I think I would like to be a professional golfer.	—.12	—.10	.66	.04	.46
7. I think I would like the work of a professional musician.	—.74	—.42	—.12	.09	.75
8. I think I could do better than most of the present politicians if I were in office.	.02	—.68	.07	.23	.52
9. I get excited very easily.	.04	.43	.54	—.04	.48
10. The average person is not able to appreciate art and music very well.	—.75	.17	—.06	—.14	.61

[a] Note: Five choices were provided for each item varying from "Strongly Agree" to "Strongly Disagree."

meaningfully about differences in over-all ability to judge others accurately. In order to do this, however, one must interpret variations in over-all accuracy as variations in the elevation of a profile (Cronbach and Gleser, 1953). This is a very different conception from variation along a single dimension, and one with important implications for future research on accuracy of interpersonal perception. At the very least, it implies that one cannot expect to adequately describe a person's accuracy of judging others with a single score. For example, if variations in over-all accuracy are conceived of as variations in profile elevations, one must also consider other characteristics of profiles such as scatter and shape.

Finally it should be noted that the scores for each item in the factor analysis were *accuracy* measures. These factors are not necessarily the same as the factors that would be obtained by analyzing the responses of the persons judged to these items. In other words, the dimensions of

TABLE XXV
ROTATED LOADINGS FOR FACTOR ANALYSIS OF INDIVIDUAL ITEM ACCURACY SCORES ON THE INTERPERSONAL RELATIONS INVENTORY

Item	Factors				
	A	B	C	D	h^2
1. When I get bored I like to stir up some excitement.	.09	.06	−.80	−.11	.66
2. I would do almost anything on a dare.	.62	−.14	−.31	.24	.56
3. I think I am stricter about right and wrong than most people.	−.32	−.51	.38	−.19	.54
4. It is all right to get around the law if you don't actually break it.	−.16	.19	.30	.60	.51
5. I often act on the spur of the moment without stopping to think.	.19	−.31	−.21	.69	.65
6. I have often gone against my parents' wishes.	.71	−.12	−.08	.06	.53
7. My home life was always happy.	−.05	.56	−.03	.15	.34
8. Before I do something, I try to consider how my friends will react to it.	.17	−.62	−.09	.17	.45
9. I seem to do things I regret more often than most people.	−.27	−.28	−.02	−.48	.38
10. Every family owes it to the city to keep their sidewalks cleared in the winter and their lawns mowed in the summer.	.67	.21	.45	−.18	.73

a Note: Five choices were provided for each item varying from "Strongly Agree" to "Strongly Disagree."

accuracy of judged religiosity are not necessarily the same as the dimensions of religiosity itself.

I. EFFECT OF SOCIAL DESIRABILITY RESPONSE SET ON JUDGING ACCURACY

A study was conducted by Cline and Richards (1964 Annual Report) on the effects of different kinds of motivation on judging accuracy where three motive strengths or incentive conditions were used with 65 Ss.

(a) *Low incentives:* The Ss were urged to do their best but no reward was provided. (b) *Medium incentive:* Subjects were told that the

most accurate judge would receive a $15 prize to be mailed out in several weeks. (c) *High incentive:* Here the subjects were told that the best judge would receive a prize of $25; the second best, $15; the third, fourth, and fifth, $10; and the sixth through tenth best judges would each receive $5. They were told that they would receive these prizes a few minutes after the experiment was completed and the money was displayed in full view at the front of the room while the experiment was taking place.

When the results yielded negligible differences in judging accuracy between the experimental groups, the researchers speculated that this might be accounted for by some uncontrolled response set affecting the judging instruments. However, in the construction of these instruments attempts had been made to control for the two most obvious response sets, namely always choosing "True" (or "False"), and always choosing the socially desirable or (undesirable) response. However, no evidence had been available to indicate whether this had been successfully accomplished or not. That this should be checked is suggested by the experience of Edwards with his Personal Preference Schedule: he thought that he had controlled for social desirability, but in fact had not (Corah *et al.,* 1958).

Accordingly, special keys were developed to give evidence about these matters.

The first step in constructing these keys was to obtain ratings of the socially desirable answer to each item of each judging instrument. These ratings were made by four persons, one Ph.D. psychologist and three advanced graduate students in psychology. The criterion for determining the socially desirable answer to an item was that at least three out of the four raters agree that the answer in question was the socially desirable one. If there was a two-two split among the raters as to the socially desirable answer, it was concluded that social desirability had been controlled on that item.

Keys were then constructed to make it possible to analyze separately all possible combinations of correct response and socially desirable response. In other words, for the Behavior Postdiction and the Opinion Prediction tests, the keys were: (a) socially desirable answer true, correct answer true; (b) socially desirable answer true, correct answer false; (c) socially desirable answer false, correct answer true; and (d) socially desirable answer false, correct answer false. For the Adjective Check, the keys were (a) socially desirable alternative correct, and (b) socially desirable alternative wrong. On each of these keys, the score is the total number correct. Another key for each of the three instruments consisted of those items where social desirability was controlled (regardless of whether the correct answers were "true" or "false"). Again the score was the total

number correct. Finally, another special key, the Popular Responses key, was constructed on the basis of the sample reported in the author's 1962 Annual Report (Cline, Richards, and Abe, 1962), who took this same judging test. A "popular response" was defined as an answer chosen by 80% of the sample regardless of whether that response was correct or not.

TABLE XXVI

CORRELATIONS AMONG VARIOUS KEYS WITHIN JUDGING INSTRUMENTS

Instrument	Key	1	2	3	4
Behavior Postdiction	1. Socially desirable answer True, correct answer True	—			
	2. Socially desirable answer True, correct answer False	—24	—		
	3. Socially desirable answer False, correct answer True	—27	27	—	
	4. Socially desirable answer False, correct answer False	39	—35	—03	—
Opinion Prediction	1. Socially desirable answer True, correct answer True	—			
	2. Socially desirable answer True, correct answer False	—38	—		
	3. Socially desirable answer False, correct answer True	—12	21	—	
	4. Socially desirable answer False, correct answer False	13	—58	01	—
Adjective Check	1. Socially desirable answer Correct	—			
	2. Socially desirable answer Wrong	—39	—		

$r \ .05 = .24$

TABLE XXVII

CORRELATIONS AMONG TOTAL CORRECT KEYS FOR THE THREE JUDGING INSTRUMENTS

	1	2	3
1. Behavior Postdiction	—		
2. Opinion Prediction	—03	—	
3. Adjective Check	14	00	—

$r \ .05 = .24$

The score on this key was the total number of popular responses chosen.

The hypothesis emerged that the Ss in making their judgments differed more in the extent to which they chose socially desirable answers than they did in accuracy. This was determined by computing correlations among the various keys within forms. If social desirability were a more important factor than accuracy, the correlation between keys where the socially desirable answer is the same and the correct answer different would be negative. For example, there would be a negative cor-

relation between the "socially desirable answer True, correct answer True" key and the "socially desirable answer True, correct answer False" key. If accuracy were more important than social desirability, these correlations would be positive.

The correlations among the various keys within instruments are presented in Table XXVI. In order to permit a further comparison of social desirability with accuracy, correlations were also computed among the Total Correct keys for the three judging instruments. These correlations are presented in Table XXVII. The results indicate quite clearly that variation with respect to tendency to choose the socially desirable response is more important in these judging instruments than variation in accuracy—at least for this study. This is true in spite of the fact that, as previously indicated, efforts had been made to control this response set. And while this is just one study, it does demonstrate the problems one encounters in attempting to develop judging instruments tailored to individual "others."

While it cannot be proved, the "social desirability" variable might also account, in part, for some of the generality found across judging instruments and "others" reported earlier in this paper.

One possible solution to this problem would be to use a paired comparison method. Use of this technique with filmed interviews stimulus material would mean that Ss would be shown separate movies of interviews of, say, two persons and then asked which of these two persons were, say, more intelligent, dominant, neurotic, etc. This would also make sense in terms of the findings of clinical *versus* statistical predictions study where the clinician-judge can best beat the actuary by ranking, ordering, or comparing several others at one time.

V. Summary

In the middle 1950's Cronbach, Gage, and several other researchers criticized naïve empiricism in interpersonal perception research. They pointed out serious weaknesses and flaws in most published studies as well as in traditional methodology common to much research in this area. The possibility of fractionating or "breaking down" global judging scores was suggested as a possible solution or way out of this problem. This would allow experimenters to obtain "pure" accuracy scores uncontaminated by other components ofttimes working at cross purposes and in contradictory ways when combined into a global index or measure. In a series of researches by Cline and Richards, the power and utility of this component approach was demonstrated and helped clarify some of the issues in the clinical *versus* statistical prediction controversy, prob-

lems with the generality of empathic ability, the judging accuracy of groups *versus* individuals, and so forth. Despite this, these components still revealed some weaknesses and under certain conditions produced conflicting and contradictory results. It was demonstrated that the nature of the scoring system could affect these components in undesirable ways and in certain situations lead to unjustified conclusions. Because of this, a new Interpersonal Accuracy (IA) measure (component) was developed which was believed free of the methodological shortcomings of the older Differential Accuracy and other components. Factor studies of this new IA component revealed that (as a "pure" measure of judging ability) it was factorially complex and not unidimensional. It involved several orthogonal factors rather than a large general factor. However, even though this was true, it was believed that one could still talk meaningfully about differences in over-all ability to judge others accurately. In order to do this, however, one must interpret variations in over-all judging accuracy as variations in elevation of a profile where scatter and shape must be considered.

It was also demonstrated that the social desirability factor exerts an extremely powerful influence in the use of all judging instruments and must be considered or reckoned with in the actual judging process as well as on the criterion development side of the problem.

The use of stereotypes in judging was found, also, to be an extremely important variable in judging. Much evidence suggested that having an accurate stereotype of a certain class or subclass of social objects or "others" being judged could, apart from any other variable, contribute significantly to accuracy.

However, it was noted at the date of writing this review that there still does not exist a fully satisfactory and validated test or procedure for measuring judging accuracy. And certainly one of the major unsolved problems is the lack of a good external criterion with which we might validate the new experimental judging tests which are occasionally developed. This is a situation not unlike that which Binet faced when the first tests of intelligence were constructed.

Because of the many methodological problems encountered where single persons or "others" are judged individually, the suggestion was made that the judging task could be greatly simplified by having one compare pairs or rank-order triads of "others" on a variety of personality variables. Thus the only discrimination required would be to indicate which of two "others" was the most dominant, intelligent, etc. Or one might rank-order three "others" on these same traits after first seeing all three films of them or observing them in interviews etc. This would solve

the problem of projection, social desirability, and idiosyncracies of the judges in their approach to standard rating instruments.

Experience has suggested that filmed interviews of "others" have to be considered as something akin to unselected items in a test. Some will prove confusing, too difficult, or too easy and will have to be discarded. And the same is true for the material in the individual judging instruments. In the author's experience a considerable amount of time and energy must be spent in refining and developing these for even a minimum payoff.

Since the author's bias is in the direction of studying "ultimate achievement" in judging, he would be of the opinion that the "best" judging measure ultimately will prove to be a sophisticated global index which will combine all the crucial elements contributing to accuracy in a measurable and meaningful way. This might involve developing five or six factorially pure tests, or a procedure similar to that used by Guilford to measure dimensions of intellect, or the much cruder Wechsler Adult Intelligence Scale where various subtest scores are combined to give a Full Scale IQ, or it might possibly even make use of a composite of a profile.

In any event, the current status of interpersonal perception research is not yet in sharp focus. A fully satisfactory measure of judging accuracy still eludes us.

References

Boneau, C. A. The effect of violations of assumptions underlying the *t* test. *Psychol. Bull.,* 1960, 57, 49-64.

Box, G. E. P. Non-normality and tests on variances. *Biometrika,* 1953, 40, 318-335.

Bronfenbrenner, V., Harding, J., and Gallwey, M. The measurement of skill in social perception. In D. C. McClelland (Ed.), *Talent and Society.* New York: Van Nostrand, 1958, pp. 29-111.

Bruner, J. S., and Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Massachusetts: Addison-Wesley, 1954.

Brunswik, E. Social perception of traits from photographs. *Psychol. Bull.,* 1945, 10, 535. (Abstr.)

Buros, O. K. *Fifth mental measurements yearbook.* New Jersey: Gryphon Press, 1959.

Campbell, D. T. An error in some demonstrations of the superior social perceptiveness of leaders. *J. abnorm. soc. Psychol.,* 1955, **51**, 694-695.

Chapin, F. S. Preliminary standardization of a social insight scale. *Amer. social. Rev.,* 1942, **7**, 214-225.

Chowdhry, K., and Newcomb, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.,* 1952, **47**, 51-57.

Cline, V. B. *The assessment of good and poor judges of personality using a stress interview and sound film technique.* Unpublished doctoral dissertation, University of California, Berkeley, 1953.

Cline, V. B. Ability to judge personality assessed with a stress interview and sound film technique. *J. abnorm. soc. Psychol.,* 1955, **50**, 183-187. (a)

Cline, V. B. *An analysis of the performance of 100 Air Force officers on the social acuity test.* USAF Pers. Train. Res. Center Tech. Memos, 1955, Project No. 7730, No. AF 18 (600)-8, Maxwell AFB. (b)

Cline, V. B., and Richards, J. M., Jr. *Variables related to accuracy of interpersonal perception.* Ann. Repts. 1958, 1959, 1960, and 1961. Group Psychology Branch, Office of Naval Research; Dept. of Psychology, Univ. of Utah, Salt Lake City.

Cline, V. B., and Richards, J. M., Jr. Accuracy of interpersonal perception—A general trait? *J. abnorm. soc. Psychol.,* 1960, **60**, 1-7.

Cline, V. B., and Richards, J. M., Jr. A comparison of individuals versus groups in judging personality. *J. appl. Psychol.,* 1961, **45**, 150-155. (a)

Cline, V. B., and Richards, J. M., Jr. The generality of accuracy of interpersonal perception. *J. abnorm. soc. Psychol.,* 1961, **62**, 446-449. (b)

Cline, V. B., and Richards, J. M., Jr. Components of accuracy of interpersonal perception scores and the clinical and statistical prediction controversy. *Psychol. Rec.,* 1962, **12**, No. 4, 373-379.

Cline, V. B., and Richards, J. M., Jr. A methodological note on Cline and Richard's studies of accuracy of interpersonal perception: A reply. *J. abnorm. soc. Psychol.,* 1963, **66**, No. 2, 195-196.

Cline, V. B., Richards, J. M., Jr., and Abe, C. *Variables related to accuracy of interpersonal perception.* Ann. Repts. 1962 and 1964. Group Psychology Branch, Office of Naval Research; Dept. of Psychology, Univer. of Utah, Salt Lake City.

Cohen, E. The methodology of Notcutt and Silva's "Knowledge of Other People": A critique. *J. abnorm. soc. Psychol.,* 1953, **48**, 155.

Corah, N. L., Feldman, M. J., Cohen, I. S., Gruen, W., Meadow, A., and Ringwall, E. A. Social desirability as a variable in the Edwards personal preference schedule. *J. consult. Psychol.,* 1958, **22**, 70-72.

Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.* 1955, **52**, 177-193.

Cronbach, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri and L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, California: Stanford Univer. Press, 1958, pp. 359-379.

Cronbach, L. J., and Gleser, G. C. Assessing similarity between profiles. *Psychol. Bull.,* 1953, **50**, 456-473.

Crow, W. J. *A methodological study of social perceptiveness.* Unpublished doctoral dissertation, Univer. of Colorado, Boulder, 1954.

Crow, W. J. The effect of training upon accuracy and variability in interpersonal perception. *J. abnorm. soc. Psychol.,* 1957, **55**, 355-359.

Crow, W. J. *Process vs. achievement and generality over objects and judging tasks.* Presented as part of 1960 APA Symposium on New Frontiers in Person Perception Research. (Privately mimeographed) 1960.

Crow, W. J., and Hammond, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.,* 1957, 54, 384-390.

Dymond, R. F., Hughes, A. S., and Raabe, V. L. Measurable changes in empathy with age. *J. consult. Psychol.,* 1952, 16, 202-206.

Estes, S. G. Judging personality from expressive behavior. *J. abnorm. soc. Psychol.,* 1938, 33, 217-236.

Estes, W. K. *The human observer and the clinician as diagnostic instruments.* Western Psychological Association Invitational Address, Seattle, Washington, 1961.

Fiedler, F. Unpublished address. American Psychological Association. Ann. Meeting, Chicago, 1960.

Flanagan, J. C. *Flanagan Aptitude Classification Test.* Chicago: Science Research Associates, 1958.

Gage, N. L. Judging interests from expressive behavior. *Psychol. Monogr.,* 1952, 66 (18), No. 350.

Gage, N. L. Accuracy of social perception and effectiveness in interpersonal relationships. *J. Personality,* 1953, 22, 128-141.

Gage, N. L., and Cronbach, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.,* 1955, 62, 411-422.

Gage, N. L., and Suci, G. Social perception and teacher-pupil relationships. *J. educ. Psychol.,* 1951, 42, 144-152.

Gage, N. L., Leavitt, G. S., and Stone, G. C. The intermediary key in the analysis of interpersonal perception. *Psychol. Bull.,* 1956, 53, 258-266.

Giedt, F. H. *Judgment of personality characteristics from brief interviews.* Unpublished Ph.D. dissertation, Univer. of California, Los Angeles, 1951.

Gough, H. G. *The opinion prediction scale.* Berkeley: Univer. of California Institute of Personality Assessment and Research, 1953.

Gough, H. G. *The assessment of social acuity.* Research Report. Berkeley: Univer. of California, 1955.

Grossman, B. A. *The measurement and determinants of interpersonal sensitivity.* Unpublished Master's thesis, Michigan State Univer., East Lansing, 1963.

Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.

Halbrower, C. C. A comparison of actuarial vs. clinical predictions to classes discriminated by the MMPI. Unpublished doctoral dissertation, Univer. of Minnesota, Minneapolis, 1955.

Hastorf, A. H., and Bender, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.,* 1952, 47, 574-576.

Hastorf, A. H., Bender, I. E., and Weintraub, D. J. The influence of response patterns on the "refined empathy score." *J. abnorm. soc. Psychol.,* 1955, 51, 341-343.

Hatch, R. S. *An evaluation of a forced choice differential accuracy approach to the measurement of supervisory empathy.* Englewood Cliffs, New Jersey: Prentice-Hall, 1962.

Holt, R. R. Clinical and statistical prediction: A reformulation and some new data. *J. abnorm. soc. Psychol.,* 1958, 56, 1-12.

Kaiser, H. F. The application of electronic computers to factor analysis. *Educ. psychol. Measmt.,* 1960, 20, 141-151.

Kerr, W. A. *Primary empathic abilities.* Chicago: Psychometric Affiliates, 1957.

Kerr, W. A., and Speroff, B. J. *Measurement of empathy* (manual). Chicago: Psychometric Affiliates, 1951.

Kerr, W. A., and Speroff, B. J. Validation and evaluation of the empathy test. *J. gen. Psychol.*, 1954, 50, 269-276.

Leytham, G. *Some factors influencing the ratings on self and others.* Unpublished Master's thesis, Univer. of California, Berkeley, 1951.

Li, J. C. R. *Introduction to statistical inference.* Ann Arbor, Michigan: Edwards, 1957.

Lorge, I., Fox., D., Davitz, J., and Brenner, M. A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychol. Bull.*, 1958, 55, 337-372.

Luft, J. *Some relationships between clinical specialization and the understanding and prediction of an individual's behavior.* Unpublished doctoral dissertation, Univer. of California, Los Angeles, 1949.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1955.

Meehl, P. E. *Clinical versus statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1955.

Meehl, P. E. Wanted—A good cookbook. *Amer. Psychologist*, 1956, 11, 263-272.

Mosier, C. I. On the reliability of a weighted composite. *Psychometrika,* 1943, 8, 161-168.

Moss, F. A., Hunt, T., and Omwake, K. T. *Social intelligence test.* Washington, D. C.: Center for Psychological Service, George Washington Univer., 1944.

Nagle, B. F. Productivity, employee attitudes and supervisory sensitivity. *Personnel Psychol.*, 1954, 18, 53-58.

OSS Assessment Staff. *Assessment of men.* New York: Rinehart, 1948.

Remmers, H. H., and File, Q. W. *How Supervise?* New York: Psychological Corp., 1943-1948.

Richards, J. M., Jr. Reconceptualization of the clinical and statistical prediction controversy in terms of components of accuracy of interpersonal perception scores. *Psychol. Repts.,* 1963, 12, 443-448.

Richards, J. M., Jr., and Cline, V. B. Accuracy components in person perception scores and the scoring system as an artifact in investigations of the generality of judging ability. *Psychol. Repts.,* 1963, 12, 363-373.

Richards, J. M., Jr., Cline, V. B., and Rardin, M. W. Stereotypes and components of accuracy of person perception scores. *J. Personality*, 1962, 30, No. 4, 601-612.

Richey, M. H. *Ability to predict responses of acquaintances: A comparison of institutional and community adolescents.* Unpublished doctoral dissertation, Univer. of Illinois, 1952.

Smith, A. E. *Rapport and social perception in mother adolescent relationships.* Unpublished doctoral dissertation, Univer. of Illinois, Chicago, 1957.

Stone, G. C., and Leavitt, G. S. *Generality of accuracy in perceiving standard persons.* Univer. of Illinois, Chicago, 1954. (Mimeo.)

Taft, R. *Some correlates of the ability to make accurate social judgments.* Unpublished doctoral dissertation, Univer. of California, Berkeley, 1950.

Taft, R. The ability to judge people. *Psych. Bull.*, 1955, 52, 1-23.

Tagiuri, R., and Petrullo, L. (Eds.) *Person perception and interpersonal behavior.* Stanford, California: Stanford Univer. Press, 1958.

Talland, G. A. The assessment of group opinions by leaders and their influence on its formation. *J. abnorm. soc. Psychol.*, 1954, 49, 431-434.

STIMULUS FUNCTIONS IN PROJECTIVE TECHNIQUES

Douglas T. Kenny[1]

HARVARD UNIVERSITY, CAMBRIDGE, MASSACHUSETTS

I. Introduction: Stimulus Factors in Projective Measures

A. VARIANCE-INVARIANCE IN PERCEPTION

The fact that stimuli facilitate and inhibit behavior is not without prior attention in the field of psychology. Ever since the early beginnings of behaviorism at the turn of this century, it has been generally assumed that behavior involves some interaction with stimulus situations. In fact, some psychologists have been so bold as to propose that such interactions constitute the empirical data of psychology. Doubtlessly here lies the basis for the protopostulate of many researchers that an intensive analysis

---

[1] On leave of absence from the Department of Psychology, University of British Columbia, for the period 1963–1965.

of the role of stimulus functions in behavior will yield findings of considerable theoretical and practical import for the discipline of psychology. Underlying most conceptions of stimulus functions is the broad proposition that the stimulus is the entering wedge into a behavior setting. The central problem or puzzle for psychology emerges when we attempt to find out what happens after an individual is stimulated.

A significant finding to emerge from early studies on the interaction between stimulation and perceiving was the clear-cut evidence that perception is not completely tied to stimulation at the sensory surface. Perhaps the classic experimental demonstrations of this lack of invariance between sensory input and perception are the phi phenomenon and reversible figures. The data on ambiguous figures immediately face us with the problem and the possibility that percepts may change without any alteration in stimulation, proximal or distal. From a historical perspective, such phenomena paved the way for the acceptance of the belief that input stimulation is only one of the many possible variables determining perception and behavior.

Such data and other interlocking findings show that we do not perceive directly either the distal stimulus or the proximal stimulus (cf. Brunswik, 1944, 1952, 1955). The former term may simply be defined as the physical object in the objective environment. The energy changes impinging on receptors refer to proximal stimulation, and they constitute the initial starting point or occasion for perception. However, if we perceive neither the distal object nor the proximal stimuli, then what do we perceive? It is at this juncture in the sequence of perceptual events that theory starts to take over. As Woodworth (1958) formulated in his characteristically middle-of-the-road fashion, it seems necessary to postulate that the individual decodes the raw proximal stimulation into objective or distal terms and then encodes the consummatory response before translating it into peripheral action. Hence, perception is the utilization of proximal stimuli as indicators of physical or social objects in the objective environment; in short, perception is a translation or decoding of proximal stimulation into distal objects.

Since the decoding of stimulation is an inferential process, veridical perception depends upon assigning the stimulation to the appropriate categories which bear correspondence to distal objects. If the inferential process decodes the proximal stimuli into an inappropriate category or code, then perception will be nonveridical, and behavior may be inappropriate. The two challenging problems in perception are to discover the empirical laws as to how proximal stimuli are decoded and how the codes are initially developed. To the extent that the formation and utilization of codes may possess individual consistencies, then these two

problems are of interest to both the perceptual theorist and the person-ologist.

B. Stimulus Functions in Early Projective Models

The foregoing oversimplified and sketchy comments on the decoding or processing of stimuli lead us to the launching platform of projective techniques. They had their origins in empirical findings dealing with the relative lack of correspondence between sensory stimulation and perception, and in individual consistencies in perceiving. The effective application of these findings to the construction of projective techniques, however, depended critically upon the early theorizing of Freud and the Gestalt psychologists concerning the role of organismic variables in determining what was perceived at the sensory surface.

By and large, the early projective test constructors and interpreters were influenced profoundly by the psychoanalytic concepts of primary and secondary processes and the role they played in coping and adapting to focal and incidental stimulation. Freud's notions on cathexes and hypercathexes of stimuli (cf. Holt, 1962) also provided a point of contact between projective test constructors and psychoanalysis. The unstructured stimuli of projective techniques were thought of as devices for reducing the reality base of stimulation, thereby permitting and encouraging primary process material to become dominant. The unclear stimuli of projective measures presumably interfere with the process of hyper-cathexis. What is equally important, the permissive test instructions and the general context of the test situation add further difficulty for the material being hypercathected. Or, putting these statements at the other end of the primary process–secondary process continuum, it was believed by early constructors of projective techniques that ambiguous stimuli, in a context of permissive instructions and test situations, weakened or impaired reality-oriented secondary process thought. Veridical perceptions would fail as a consequence of difficulties in the hypercathexis of unstructured stimuli.

However, the veridicality of perception also had to be dealt with by projective personologists. While Rorschach provided the penetrating insight that the manner of perceiving was a function of intrapsychic functions, he did not attempt to develop any systematic theory to explicate his perceptual findings. This conceptual task was left to others. Unfortunately, as Lindzey (1961) has shown, this assignment is still far from completed.

Endeavors to explain the stability of perceptions in projective tests have most frequently invoked structural or autochthonous principles of perceptual determination. Extrapolation of gestalt principles of per-

ceptual organization has been primarily limited to the Rorschach test (Arnheim, 1951, 1954; Beck, 1933; Belden and Baughman, 1954; Berliner, 1955; Brosin and Fromm, 1942; Klein and Arnheim, 1953; Rickers-Ovsiankina, 1960; Stein, 1951; and Wertheimer, 1957). Despite the fact that many authors have proposed that projective techniques have affinities to gestalt principles, the application of these principles to such tests is by no means clear. The main shortcoming of the theorizing relating principles of perceptual organization to projective techniques is that it has been very casual and has not led to much fruitful experimentation.

The essentials of a gestalt explanation of projective percepts, based in large part upon Wertheimer's (1957) discussion of the relations between the Rorschach and gestalt laws of organization, starts with the notion that a projective stimulus is a multiple reversible figure, with the autochthonous factors setting limits on the influence of intrapsychic factors. If the autochthonous variables are excessively strong, then personality parameters will play only a negligible part in the formation of percepts and cognitions. On the other hand, stimuli that are weak in structural determinants (ambiguous stimuli) will maximize personality processes in perception and cognition. In spite of ambiguous test material, however, the principles of perceptual organization would still have their effects on projective responses. The specific consequences would vary within a particular projective technique, especially for any test with test items varying in degree of stimulus ambiguity.

Within these two boundary conditions, the gestalt principles of perceptual stabilization have been assumed to operate as follows:

1. *Law of good continuation.* Perceptual structuring of projective stimuli will eliminate the irregularities in the stimulus material. For example, an irregular round object on the Rorschach may be perceived as a bowling ball, just as small parts may be perceptually omitted to provide a good gestalt.

2. *Law of similarity.* Similar stimuli will arouse perceptual processes leading them to be perceptually united or grouped. It is probable that similar forms and similar colors are the two most important similarity factors which lead to perceptual separation of the Rorschach cards into major and minor details.

3. *Law of closure.* Incomplete stimuli will create perceptual processes that will induce complete figures. For example, contour gaps on the Rorschach should elicit absolute or partial completion, thereby creating unitary percepts. Difficulties in closure may also appear on the Rorschach, e.g., the incomplete closure at the top of Card VII.

4. *Law of proximity.* Percepts will be formed between structural parts of a projective stimulus that are near one another. This principle,

in conjunction with that of symmetry, helps to divide the inkblots into their major details. The interaction of these two principles is seen most clearly at work on Card X of the Rorschach, where whole responses are difficult to form.

5. *Law of symmetry*. Symmetrical aspects of the projective stimulus will facilitate perceptual grouping. Symmetry of shape undoubtedly plays a critical role in form accuracy on the Rorschach.

While these structural laws of perception are assumed to apply to all individuals and to all stimuli, it may be hypothesized that their relative influence will vary with the structuredness of the stimulus material. From this assumption, it follows that inkblot stimuli weak in autochthonous factors would increase the frequency of color-form percepts and vague whole responses and decrease form plus percepts and popular responses. These predictions rest on the implicit assumption that form dominant *prägnant* percepts reflect autochthonous grouping principles. The testing of these predictions would necessitate the construction of inkblots that systematically vary in structure. This formulation bears close affinity to the recent analysis of Rorschach protocols, based on Werner's (1957) developmental stage theory of cognition, first applied by Friedman (1953) and summarized by Hemmendinger (1953, 1960).

It is to be noted that the gestalt and psychoanalytic models of stimulus functions have in common the conjecture that ambiguous stimuli heighten the influence of intrapersonal variables in perception.

Given the empirical findings on ambiguous stimuli and the theoretical formulations of psychoanalysis, it is not at all surprising that Rorschach (1921) and Christiana Morgan and Henry Murray (1935) chose unclear visual stimuli to maximize the variance between the physical stimulus and response output. Almost inevitably, the Rorschach and the Thematic Apperception Test (TAT) became viewed as open-end functions which tapped the major motivational, cognitive, and defensive systems of individuals. They are open-end stimulus functions in the sense that any particular ambiguous stimulus is open to multiple perceptions and interpretations. Of course, such tests could equally well have been called open-end or free-response assessment techniques to the extent that the subject has to produce his own responses.

From the working assumption that ambiguous situations facilitated the operation of drive-oriented thinking, it was further assumed by projective theorists that the greater the ambiguity, the greater the personality revelation. That is, the somewhat cavalier assumption was made of a monotonic relationship between stimulus ambiguity and personality revelation. With respect to fantasy productions, McClelland (1958, p. 31) has put this assumption rather succinctly in his statement that fantasy

"capitalizes on the generalization, repeatedly confirmed in studies of the effects of motivation on behavior, that the more ambiguous the situation, the greater the effect of motivation." This hypothesis is consonant with gestalt predictions, for Wertheimer (1957, p. 212) also posits, "With more ambiguous material, a greater influence of the personality factors can be expected." We shall cite evidence later on to show that this assumption may have a questionable empirical base.

The foregoing discussion presents the stimulus model guiding early theorists and interpreters of projective techniques. Naturally the question arises as to whether this stimulus model is still adhered to in contemporary times. Indirect evidence of the continued acceptance of this model is shown by the results of a recent study by Seeman and Marks (1962). Thirty-six experienced Ph.D. clinical psychologists were asked to rank order tests on the dimensions of stimulus ambiguity, response freedom (degree of restriction of the response), and "depth" (availability to consciousness). The tests were the Bender-Gestalt, Rorschach, Sentence Completion, Strong, Szondi, TAT, and Word Association. The rank order correlation between stimulus ambiguity and depth was .86, reflecting the continued adherence to the proposition that the greater the stimulus ambiguity, the greater the depth or revealingness. The correlation between response freedom and stimulus ambiguity was .91. Thus the concept of stimulus ambiguity seems to imply the conception of many degrees of freedom in response production. Interestingly enough, response freedom and depth evidently reflect the same dimensions in the thinking of clinical psychologists since the rank order correlation between the two variables was 1.00.

With a projective model of this type, based largely on gestalt and analytic principles, it is not surprising that the specific role of the stimulus in evoking responses was largely overlooked by early theorists. Content and formal analysis of projective protocol data became *personality centered*, with the role of how the stimulus articulates with inner psychological systems ignored. In a singularly lucid and picturesque manner Rosenzweig (1951, p. 215) summed up this position when he said "The stimulus was dethroned. . . ." The stimulus remained in limbo until approximately 1950. An analysis of the historical reasons for the stimulus remaining in neglected confinement has been presented elsewhere (Kenny, 1961).

C. Recent Developments Focusing Attention on Stimulus Functions

The past decade or so has witnessed a gradual renaissance of theoretical and empirical interest in the role that stimuli play in prompting

responses to projective techniques, especially for thematic apperceptive methods. What factors contributed to the restoration of the stimulus to grace?

Certain developments within psychology as a whole undoubtedly provided the strongest impetus to a fresh interest in the stimulus. Two, somewhat correlated, developments should be singled out for special mention. In the first instance, the examination of the behavioral determinants in perception, in contrast to the autochthonous factors, helped to rekindle an interest in the perceptual aspects of projective techniques. Bartlett's (1932) and Sherif's (1935) studies on autism set the stage for this research trend, as did the pioneering studies of Sanford (1936, 1937) and Levine, Chein, and Murphy (1942) on the influence of the hunger drive upon the interpretation of ambiguous figures. The broad assumption that needs influence perception was given close attention by the studies of Bruner and Postman and their colleagues (e.g., Bruner, 1951; Bruner and Postman, 1947; Bruner and Postman, 1949; Postman, Bruner, and McGinnies, 1948; Postman and Solomon, 1949) and a legion of other investigators.

The studies on perceptual defense and vigilance soon came under a bombardment of criticism, especially for failing to control adequately the nature of the stimulus material. Two of these attacks centered around the necessity of controlling the relative frequency of word usage (Solomon and Howes, 1951) and of ensuring the emotional relevance of the stimuli to the subjects (Chodorkoff, 1955).

That the stimulus properties in such experiments are intimately associated with the resulting perceptual reactions was hinted at by one of the earliest studies in this broad area of investigation. The results of Levine, Chein, and Murphy (1942) suggested that the more ambiguous stimuli favored reality processes, whereas the less ambiguous figures facilitated the operation of autistic processes. Not only did this experiment demonstrate the need for careful manipulation of stimulus material, but it also questioned the validity of the assumption that the influence of need increases with increasing stimulus ambiguity.

The relevancy of these experiments to projective techniques was obvious, in that they helped to make clear the fact that any interpretation of projective material in terms of personality dynamics alone would be grossly deficient. The data showed unequivocally that the findings were partly dependent upon the stimulus.

Further developments within psychology also helped personologists to look again to the stimulus, namely, the new theorizing on perception (e.g., Bruner, 1948, 1957) and the advances in psychoanalytic ego psychology (e.g., Hartmann, 1950, 1958). This new theorizing indicated

that a "new" look had to be given to the role that stimulus variables might play in perception and in projective methods.

Thus developments within psychology as a whole contributed to a renewed appreciation of the importance of stimulus functions in behavior determination. The convergence of this trend with two developments within projective techniques themselves helped to forge a sharpened interest in the stimuli of these instruments. On the one hand, the research journals in the early 1950's were honeycombed with evidence of the failures of the Rorschach and the TAT in personality assessment. Kelly (1954), reflecting the growing belief that these techniques lacked validity, suggested that social psychologists might wish to study the adherence to tests in the face of so much negative validity. (For the clinical psychologist, projective techniques are like theories to many experimental psychologists; they will not be given up until supplanted by more valid techniques.) The problems encountered in attempting to establish the validity of the two standard projective instruments led many investigators to attempt construction of their own custom-built projective tests. The latter instruments immediately caused the researcher to look at the stimulus, and, as a consequence, much new evidence was produced about stimulus functions in projective test material. Some of this provocative research will be examined in a later section. On the other hand, a small number of investigators had always been interested in determining the role that the stimulus played in projective instruments. Some investigators, for example, were examining the problem of whether or not "color shock" on the Rorschach is attributable to the effects of color by removing the color from the chromatic cards. Other investigators were concerned with the issue of stimulus ambiguity and personality revelation in that TAT.

The coalescence of these trends within psychology as a whole with those in the projective area has recently prompted many investigators to look more closely at the stimulus aspect or input of projective techniques. The results of these studies will ultimately be useful in the construction of better personality assessment techniqus and will also help clinical psychologists and others to understand more fully how projective instruments function psychologically.

II. A Set of Stimulus Postulates for Projective Methods

As the preceding discussion has shown, early attempts to specify the stimulus model of projective techniques usually involved weak analogizing from psychoanalytic and gestalt theory. Many assumptions about the stimulus were implied, but not explicitly stated. We shall now try to

make these assumptions more explicit. A set of stimulus postulates will be developed, from which it is hoped that specific predictions may be derived, and tested. The postulate set requires verification.

The first systematic endeavor to specify clearly the interpretative assumptions, including those relating to the stimulus, was made by Lindzey (1952) for the TAT. Similar, but more modest, attempts have been provided by Henry (1956), Kagan (1960), and Murstein (1963b). These beginnings were expanded upon by Lindzey (1961) for all projective techniques in his book *Projective Techniques and Cross-Cultural Research*. However, since these endeavors have glossed over stimulus assumptions somewhat, our purpose is to bring into sharper focus the stimulus postulates that seem to be reasonable in terms of theory and empirical evidence.

As a starting point, the protopostulate for the entire set is: *Responses produced to projective techniques are controlled, to varying degrees, by by the projective stimulus.*

Researchers and practicing clinical psychologists have pretty well abandoned the earlier notion that projective techniques stand majestically above the influence of stimulus variables, revealing in pristine purity true personality variance. Three major kinds of empirical data have helped to fashion the proposition that responses to projective techniques are deeply dependent upon the stimulus. Since most of the important research will be discussed in later sections of this paper, no attempt is made to provide a detailed discussion of it at this point. The diverse kinds of data indicating that the stimulus, in part, controls the response to projective stimuli are as follows:

1. Normative Data

Popular responses on the Rorschach indicate that specific areas of this test are well structured, at least for certain age, sex, and cultural groups. Moreover, the fact that only 10% of the percepts to the Rorschach are "man-made" objects may be attributed to the inkblot stimuli (Edmonston and Griffith, 1958). The general shape and form of the inkblots restrict the type of response that may be made to the Rorschach. Similarly, common themes to particular TAT cards (Cox and Sargent, 1950; Eron, 1948, 1950; Lindzey and Goldberg, 1953; Rosenzweig and Fleming, 1949; Wittenborn, 1949) support the proposition that responses to thematic apperception techniques are influenced significantly by the stimulus. No useful purpose would be served by reviewing normative data in this paper. We shall indicate briefly some of their implications.

One of the most important implications of normative data is that clinical psychologists must eliminate from the total test variance that part

due to the stimulus properties of their instruments. Research is required on how best to do this. At the moment, clinicians must perform this delicate task in a highly subjective manner. A related implication of normative data is that an individual's level of psychological functioning will vary within an instrument as a consequence of the varying stimulus pull of the projective test material. That is to say, it would be unwise to interpret all data from a single test as if it were originating from a single level of personality functioning.

The last item deserves more attention than in has received in either research or the clinical application of projective tests. Projective personologists have frequently made the assumption that similar responses represent similar intraindividual dynamics. However, the data on common themes and popular responses suggest that such an assumption is very questionable. A more reasonable assumption is that intraindividual consistency in the same class of behavior due to functionally similar test stimuli reflects a different kind of psychological functioning than does intraindividual consistency in the same class of behavior to dissimilar test stimuli. This supposition immediately leads one to the representativeness and type of stimuli that should be used within a projective test. This topic is discussed under the first postulate.

2. Connotative Meanings of the Stimuli

The connotative meaning of the Rorschach inkblots has received a great deal of attention recently, with the result that it is now clear that there is a great deal of communality in the meanings assigned to specific inkblots. Findings from this research show that the percepts to inkblots are, in part, anchored to the stimulus properties of the inkblots.

Historically speaking, the recent rash of studies on the connotative meanings of inkblots stems from the hypothesis that specific Rorschach inkblots have special symbolic meanings. Card IV, for example, is the so-called "father" card, Card VI is the "sex" card, and Card VII indicates "mother." In early research (e.g., Meer and Singer, 1950; Taniguchi, De Vos and Murakami, 1958) on the mother-father cards, subjects (Ss) were asked to indicate the card that reminded them of father and the card that reminded them of mother. Today, however, Osgood's semantic differential (Osgood, Suci, Tannenbaum, 1957) has become the popular device to investigate not only the father-mother hypothesis but also other stimulus values of the Rorschach (Little, 1959; Loiselle and Kleinschmidt, 1963; Rabin, 1959; Rosen, 1960; Zax and Benham, 1961; Zax and Loiselle, 1960a,b; Zax, Loiselle, and Karras, 1960). The semantic differential has also been used to assess the stimulus properties of the Blacky Picture Test (Stricker, 1963) and of the Bender-Gestalt (Tolar, 1960).

The most striking contribution that the semantic differential studies have made is the general finding that each Rorschach inkblot has a somewhat unique meaning. A consensus of the findings on the semantic differential research and the specific studies (Levy, 1958; Mayer and Binz, 1961; Meer and Singer, 1950; Rosen, 1951; Sappenfield, 1961; Schleifer and Hire, 1960; Taniguchi *et al.*, 1958) testing the mother and father notions permits the following characterizations of the Rorschach inkblots:

Card I: ugly, dirty, cruel, large, strong, rugged, active, ferocious
Card II: happy, strong, active, fast
Card III: good, clean, happy, light, active, fast
Card IV: bad, dirty, cruel, strong, heavy, slow, ferocious, masculine
Card V: light, active
Card VI: large
Card VII: good, beautiful, clean, happy, light, delicate, peaceful, feminine
Card VIII: clean, active
Card IX: strong, active, hot
Card X: good, beautiful, clean, happy, light, active, fast

While the foregoing data show that individual inkblots tend to instigate somewhat specific connotative reactions, they do not, however, indicate what specific properties of the inkblots determine the reactions.

3. *Experimental Manipulations*

Much attention has been given to the general problem of how projective responses change as a function of experimental manipulation of the projective stimulus. Researchers (e.g., Framo, 1952; Horiuchi, 1961; Stein, 1949) have impoverished Rorschach inkblots physically by tachistoscopic techniques. They have shown generally that the form quality of whole responses (W) and percentages of usual detail responses (D) increase with longer exposure times. Another kind of stimulus manipulation of the Rorschach has involved alteration of some assumed critical property of the inkblots, such as color. Since this type of study will be reviewed in depth later, it suffices to say at this point that such stimulus changes do produce some shifts in the response.

Similarly, manipulation of the stimulus properties of thematic apperception pictures has received a great deal of attention. Several workers have examined fantasy production as a function of similarity between the central figure in a picture and the subject. Investigators have also tested the hypothesis of a monotonic relationship between stimulus ambiguity and personality revelation. Several recent studies have focused on the problem of what kind of picture is most likely to reveal drive manifesta-

tions. These and other tantalizing questions will be examined in later sections. Accordingly, it need only be stated that the research evidence shows that there are important interactions between projective test performance and stimulus properties of the pictures.

At this juncture in the development of projective techniques, what is required is a more precise delineation of these interactions. An attempt will be made in this paper to present a framework within which the interactions between stimulus properties of projective materials and response performance can be intensively explored. In essence, then, the following set of postulates attempts to provide this framework.

A. POSTULATE ONE: PSYCHOLOGICAL PROCESSES AS A FUNCTION OF TEST STIMULUS MODEL

Similar responses to a projective test will increasingly reflect analogous dynamics as a test approximates a cumulative homogeneity test stimulus model.

Following the views expressed by Fiske (1963), three test stimuli models may be distinguished in current projective methods. The commonest model today is the *pure relative frequency* one and is best exemplified by the Rorschach and the TAT. Other perceptual tests based on this model are the Behn–Rorschach (Zulliger, 1946), the Holtzman-Inkblot Technique (Holtzman, Thorpe, Swartz, and Herron, 1961), and the Z-test (Zulliger, 1948). The Picture Story Test (Symonds, 1949), the Michigan Picture Test (Andrew, Walton, Hartwell, and Hutt, 1951; Andrew, Hartwell, Hutt, and Walton, 1953), the Make-A-Picture Story Test (Shneidman, 1948), the Four Picture Test (Lennep, 1948), the Children's Apperception Test (CAT) (Bellak, 1954), and the Object Relations Test (Phillipson, 1955) may serve as additional examples for constructive projective techniques which employ the pure relative frequency model.

In this model, many different stimuli are used to assess numerous personality characteristics and an inference about a particular personality characteristic is formed by counting similar response productions from relatively different test stimuli. For example, on the Rorschach one counts the number of whole responses, major detail responses, minor detail responses, and the like. On the TAT, we count the number of aggressive themes, succorance themes, achievement themes, and so on. It is to be noticed that the same stimuli are used to assess different personality variables and, as a consequence, the counting operation is performed independent of the stimuli.

A second and increasingly popular approach with projective methods is to custom-build tests based on the specific relative frequency model. With this model the investigator selects stimuli for the measurement of

a specific personality characteristic, and simply counts the number of times that the characteristic manifests itself in the response productions. Unlike the next model to be described, no attempt is made to order the stimuli in terms of their evoking strength for the particular personality variable under consideration. In principle, any stimulus which seems relevant to the personality dimension under investigation will do. The general usefulness of this model was highlighted by McClelland (McClelland, Atkinson, Clark and Lowell, 1953) and his colleagues, who constructed TAT-type pictures specifically designed to engage the achievement motive. Kagan (1956) made use of this test model when he constructed a special set of TAT pictures to measure aggression toward peers. The Levy Movement cards (Rust, 1948) is a perceptual test employing the relative frequency model. The use of doll play with structured questions around sibling rivalry also employs this model (Levy, 1937).

A third, but still relatively uncommon, approach with projective techniques is the cumulative homogeneity model. The two distinguishing features of this model are that the stimuli are ordered by scalogram analysis along a continuum of decreasing ambiguity with which they measure a single personality characteristic. Since this model assumes that the stimuli have exactly the same order for all subjects, any two subjects with the same total score should have identical answer patterns. For projective techniques, the stimuli would be ordered in terms of increasing percentage of individuals manifesting a given variable, such that once an individual gives a response reflective of the variable, he should continue to give similar responses to the remaining stimuli. Thus the stimuli are arranged in order of increasing stimulus demand for a variable. A subject with a weak tendency for a variable would have to continue up the scale of decreasing ambiguity much further than would a person with either a moderate or strong tendency. Preferably, a rectangular distribution of stimulus demand frequencies should obtain for such a test, with a minimum of stimulus items of demand frequencies near .5 since such items are unstable (Fiske, 1963).

Guttman's (1950) scalogram analysis was successfully applied by Lesser (1958) in the construction of 10 pictures to measure fantasy aggression. This study provides convincing evidence that the cumulative homogeneity model may be applied to projective techniques, providing that the test constructor designs his materials to ensure the elimination of stimulus elements that may evoke unwanted response systems. Auld, Eron, and Laffal (1955) were only partially successful in their application of Guttman's scaling method to the TAT because the pictures were not purified of unwanted stimulus elements. A method basically similar to scalogram analysis, but dealing more directly with the concept of a thresh-

old, has been employed by Barron (1955) to inkblots. In fact, Barron's achromatic inkblots, constructed specifically to provide a threshold measure of human movement (M), would lend themselves quite readily to a scalogram analysis. A study of Spivak, Levine, and Sprigle (1958) with Barron's series of blots show that the percentage of subjects giving M to the blots ranged from .00 to .72, with an average increment of .03 between adjacent inkblots. Moreover, their data indicate that a good approximation to a rectangular distribution could be obtained with only 13 of the original 26 inkblots, with an average step interval of .05.

After this brief examination of models underlying various kinds of projective methods, we may now return to the original proposition that similar responses to a projective test will increasingly reflect analogous dynamics as the test approximates a cumulative homogeneity model. This postulate explicitly asserts that the pure relative frequency and the cumulative homogeneity models represent polar opposites on a continuum of internal consistency, with the specific relative frequency model in the middle.

The published data on projective techniques based on the pure relative frequency model indicate that prudence must be shown before superficially similar responses are combined to reflect similar dynamics [see Jensen (1959) for a comprehensive review of the literature on the internal consistency of standard projective methods]. The vital importance of cautiousness is indicated by the results of a rarely cited study by Wittenborn (1950). He examined the hypothesis that responses to the TAT which are similar to each other should be more frequently related to each other than should dissimilar responses. Three separate tests of this hypothesis were made. In the first analysis, he tested for the interrelationships between the "good woman" response role assigned to the woman in cards 4, 6GF, and 13MF. As none of these relationships were significant, the "good woman" role must reflect different dynamics in the three cards. The second analysis tested for significant interrelationships between "conflict evidence" frequently ascribed to the heroes for cards 2, 4, 7BM, and 13MF. Again, the conflict response categories were not significantly interrelated. The third analysis tested for the interrelations among "hostile" responses given to cards 7BM, 7GF, 4, 6BM, 6GF, and 13MF. The data also failed to support the assumption that hostile feelings reported to different TAT cards will be significantly related to each other.

Confirmation of Wittenborn's early finding that similar responses to different TAT cards reflect dissimilar dynamics is contained in the studies by Child, Frank, and Storm (1956), Lindzey and Goldberg (1953), and Epstein (1964). Employing a group-administered TAT, Child, Frank, and Storm (1956) found that internal consistency estimates for 10 need

variables ranged from −.07 to .34, with a mean of only .13. Similarly, Lindzey and Herman (1955) obtained split-half reliability figures ranging from .12 to .45, with 8 TAT cards on six different personality variables. In a recent study (Epstein, 1964), conducted under the author's guidance, a rank order method was first used to order nine TAT pictures in terms of hostility. The initial nine cards (1, 2, 3BM, 4, 6BM, 11, 14, 18BM, 18GF) were selected on the basis of experience to represent a range of aggression in terms of card pull. Cards 2, 4, and 6BM were eliminated from further study because there was considerable variation between subjects in the ranks assigned to them. Cards 1 and 14 were designated the low-aggressive, cards 3BM and 11 as moderately aggressive, and cards 18BM and 18GF as highly aggressive. Previous research and clinical experience would seem to support the classification of these six TAT cards. Forty volunteers from an introductory course in psychology told TAT stories to these cards. The stories were scored on Stone's (1956) aggressive content scale, Purcell's (1956) internally based punishment score, and his externally based punishment score. Internal consistency was assessed by intercorrelating the data on each of the three sets of aggressive cards for the three aggressive measures. Of the nine correlations, only one was significantly different from zero. The correlation between the low and the medium aggressive cards on internal punishment was .48 ($p < .05$). The results of this study provide further support for the proposition that superficially homogeneous responses to projective tests based on the pure relative frequency model may not be assessing similar psychological dynamics.

While the first postulate leads to the prediction that, in general, a projective test composed of homogeneous stimuli will be more sensitive to analogous psychological processes, direct empirical evidence is lacking on this prediction for projective tests based on either the specific relative frequency model or the cumulative frequency model. However, this assumption has received some highly indirect empirical support from a study by Carr (1956) on intraindividual consistency in response to projective tests of varying degrees of ambiguity. The major finding of this research was that there was greatest consistency in level of inference about affect between the TAT and Rorschach and between the TAT and Sentence Completion Test. If the Rorschach and Sentence Completion test stimuli are thought of as extremes on a continuum of stimulus homogeneity, then the finding of greatest consistency between each of these projective techniques and the TAT lends indirect support for our assumption. This study also gives no support for the hypothesis of an inverse relationship between psychological processes as a function of test stimulus dimensions. In other words, irrespective of the kind of stimuli employed, test indices of apparently similar psychological processes will

either produce positive relationships or no appreciable correlations. The correlations will not be negative, as some individuals have hypothesized.

Summary. Despite the lack of direct evidence bearing on the first postulate, there is reason to believe that it will hold up because of the test construction procedures implied by the specific relative frequency model and the cumulative frequency model. Furthermore, theory and research in the fields of aptitude and attitude measurement point to the reasonableness of the postulate.

B. POSTULATE TWO: FUNCTIONAL STIMULUS

Valid interindividual and group comparisons necessitate the assumption of constancy of the functional stimulus on a projective technique for all individuals.

Projective personologists have been slow in their acceptance of this postulate, perhaps because such a proposition severely complicates no-end clinical inferences and research. Research evidence in many different areas of psychology has gradually taught investigators at least one important lesson about the stimulus in psychological experiments. It is that a subject may not be responding to the stimuli that the experimenter thinks he is. Hull (1943) was aware of this problem when he distinguished between the potential stimuli of a situation and the actual stimuli received by a subject. In a somewhat similar fashion, Underwood (1963), in a lucid discussion of stimulus selection in verbal learning, has distinguished between the apparent or *nominal* stimulus and the actual or *functional* stimulus. The nominal stimulus is what the experimenter presents to the subject, whereas, the functional stimulus refers, in part, to the properties of the nominal stimulus which the subject employs to cue his behavior. In many psychological experiments, there may be serious discrepancies between the nominal and functional stimulus, rendering any isomorphic assumption between the nominal and functional stimulus suspect.

With respect to projective techniques, the nominal stimulus is what the psychologist presents to the individual. For example, the item stem "My mother . . . ," the scene portrayed in picture 6BM of the TAT, and Card VII of the Rorschach are nominal stimuli. For these materials the term *functional stimuli* refers to how an individual perceives the meaning of the item stem; how he perceives what is portrayed in picture 6BM; and what meaning he attributes to card VII of the Rorschach. Do each of these three projective items signify "mother" as a functional stimulus for most Ss? The information necessary to answer this query would be shown by the degree of consistency of perceptual reactions to these projective stimuli. That is, a functional stimulus is specified by the perceptual

decoding of the nominal stimulus by the individual. A simple physical specification of the nominal stimulus will not enable the investigator to discover the functional stimulus. The notion of a functional stimulus is expressed by the following proposition:

$$FS = f(Dimensions\ of\ nominal\ stimulus \leftrightarrow decoding\ schemes)$$

where the double arrow indicates that the stimulus dimensions of the nominal stimulus and the decoding systems of an individual interact continuously with one another.

The assumption that the FS must be the same for all individuals, if valid individual or group comparisons are to be made, may be brought into initial focus by dealing first with the problem of cross-cultural comparisons with projective techniques. Suppose we wish to compare n-Achievement between the Wisconsin Menomini and Algerian Arabs by means of a TAT-type test. Before constructing his test pictures, the projective personologist would obtain ethnographic data on the kinds of cues in each society which arouse achievement-oriented dispositions. Successful construction of one set of pictures, suitable for both groups, depends upon discovering trans-cultural cues for achievement arousal. If the nominal picture stimuli do not possess trans-cultural equivalence or functional stimulus equivalence for both groups, then group differences between the Arabs and Monomini could not be attributed to differences in achievement drive. That is, differential group perceptions of what is portrayed in the pictures would vitiate any valid cross-cultural comparisons. On the other hand, the investigator may construct different sets of pictures for the separate groups. In principle, such a procedure is entirely suitable, providing his ethnographic data permit him to match the cue strength for achievement in the two sets of pictures. Of course, the practical problems involved in such an approach are almost insurmountable, largely because of the absence of good ethnographic data which would enable the investigator to equate different nominal stimuli for functional stimulus equivalence.

Most, if not all, projective personologists would readily grant the need for functional stimulus equivalence in a comparison of two widely different cultural groups. However, the whole thrust of the principle at issue has not been so widely accepted in the comparison of various subgroups within the North American culture. It would be most unfortunate if the general principle of functional equivalence of stimulus meaning were disregarded when group comparisons were performed between groups within the same society, whether the comparisons be between males and females, young and old subjects, lower and middle class groups, or other similar group comparisons.

Consider, for example, the following problem: Suppose we wished to compare the achievement drive of Radcliffe students and home economics, students from a state university. It probably would be inappropriate to use a picture which presents a scene of a young lady baking a cake as a means of assessing the achievement drive differences between the two groups. It is most unlikely that such a scene would possess equal cue strength for achievement or similarity in perceptual meaning for the two groups. Likewise, Card 1 of the TAT probably would not be suitable in any valid comparisons of differences in the achievement drive for young and old persons, largely because older persons usually do not perceive this card as an achievement-arousing scene (Atkinson, 1950).

What is equally important, the second postulate applies not only to group comparisons but also to comparisons between individuals. It is not possible to assert that one individual has more of one personality variable than another person, unless one can reasonably assume that both individuals decoded the nominal stimulus into the same functional stimulus. That is, valid individual comparisons require the assumption that the individuals concerned agree on what they perceive in the projective stimulus. If one says it is an aggressive stimulus and the other says it is an achievement stimulus, then one is not in a position to say on the basis of TAT stories that one individual has more achievement or aggression than the other person.

Until recent times, very little attention has been given to the problem of ensuring that the projective stimuli have equivalent functional properties before two or more individuals or groups are compared. Moreover, research aimed at the construction of projective stimuli which have similar functional stimulus properties for different groups should make a significant contribution to the personality measurement field.

Numerous studies (Lindzey, 1961) have utilized projective techniques in cross-cultural research. While it is not surprising that these studies show many group discriminations, the exact meaning of the differences remains a puzzling issue. Within the context of this paper, a basic factor producing the obtained differences may be attributable to differential perceptions of the meaning of the test materials. For example, the failure of a group on the Rorschach to perceive the central detail on Card I as human, the figures in Card III as human, and the whole blot of Card V as a winged animal may only reflect cross-cultural differences in perception. The form, shading, figure-ground relationships, color, and textural aspects of the inkblots readily lend themselves to "perceptual biases" for divergent cultural groups. Cross-cultural differences may merely indicate "perceptual biases" and not genuine differences in personality. As far as this author is aware, no attention has been given in cross-cultural research

to the problem of determining if individuals from different societies can actually perceive certain percepts within specifically designated areas of the Rorschach. Until such data are obtained, there is no need to assume that cross-cultural differences on the Rorschach originate in personality differences.

Similarly, social class and age differences on the Rorschach for North American samples may only reflect differential "perceptual biases" and not differential personality functioning. While research (Auld, 1952; Haase, 1956; Riessman and Miller, 1958; Stone and Fiedler, 1956) indicates that there are social class differences on the Rorschach, it remains to be shown that these discriminations can be ascribed to personality variables. Similarly, studies (Douvan, 1958; McArthur, 1955; Veroff, Atkinson, Feld and Guria, 1960) with the TAT indicate that there are differences in need-Achievement between the middle and lower classes. The differences may be class-oriented "perceptual biases" transcending personality factors. We are again faced with the need for research to determine if members of different social classes can perceive percepts in specifically marked-out areas of the Rorschach. Moreover, the same matters are at issue in age and sex comparisons on the Rorschach and TAT.

As part of a comprehensive research program on stimulus factors in thematic apperceptive techniques, this author and his associates have conducted a series of studies to determine if TAT cards have functional stimulus meaning for different groups. Underlying the methodology of these studies is the premise that when a projective stimulus is presented to a subject he will attempt to identify, categorize, or otherwise decode the nominal physical dimensions of the stimulus. Such decoding systems are the operational specifications of what we mean by functional stimuli. If a person is asked to describe what he sees in card 3BM of the TAT, he may say, "This is a person who is broken up over something," or "The person is thinking of or has tried to take his own life." Such reactions are attempts to assign the nominal stimuli of a TAT card to decoding systems. They represent what Bartlett calls "an effort after meaning" (1932, p. 44). While stimulus oriented, most individuals clearly go beyond the nominal physical stimulus properties of TAT cards when they are asked to describe what they see in each card. By decoding the nominal stimulus, the perceiver adds something to the nominal stimulus, thereby making the functional stimulus (decoding system) somewhat different from the nominal stimulus.

The first phase of this research required Ss to examine each TAT card for a 20-second period and to "describe what is in each picture." The next step was similar to a critical incident study in that the card

descriptions were assigned to decoding categories that seemed best to reflect the codes used by the subjects in assigning functional stimulus properties to the TAT pictures. That is, the decoding systems were not formed *a priori*, but were empirically derived from a detailed analysis of the perceptual reactions. In the first studies (Kenny, Harvey, and Wilson, 1961), a total of 40 hospitalized psychiatric patients and 40 normal control Ss were used to obtain the initial decoding reactions to 26 TAT cards—1, 2, 3BM, 3GF, 4, 5, 6BM, 7BM, 8BM, 9GF, 10, 11, 12M, 12BG, 13MF, 14, 15, 17BM, 17GF, 18BM, 18GF, 19, and 20. The Ss were told that the experimenter was interested in having people tell her what they see in pictures and that during a 20-second exposure of each card they were to describe what they saw in each picture. The patients (23 male and 17 female) were severely disturbed psychoneurotics who were institutionalized in a provincial mental hospital, in which the maximum allowable period of hospitalization is 4 months. The controls or normals were matched with the patients for level of education. The modal number of perceptual codes, required to categorize the reactions of the Ss per card, was 8, with a range of 6 to 12. Scorer reliability, based on two independent judges assigning the perceptual reactions to the codes, was 91.2% agreement.

Do severely disturbed neurotics and normals employ different decoding mechanisms in assigning functional stimulus meaning to TAT cards? Out of a total of 216 decoding categories, only 19 showed significant chi square tests at the .01 or .05 level between the groups. Accordingly, it may be concluded that the functional stimulus meaning of TAT cards for normal and neurotic adults manifests high communality.

The propitious finding of a striking congruence between normals and neurotics in their decoding systems should not be generalized to other group comparisons. In order to determine whether extreme cultural differences make for dissimilar decoding reactions to the TAT, another study (Moriya, 1962) compared the decoding mechanisms of Japanese and Canadian samples. On the basis of data obtained from previous studies in this series, the five most common perceptual coding reactions for each TAT card were prepared and the Ss were asked to check the perceptual reaction which they thought best described what they saw in each TAT card. The Japanese sample consisted of 40 male and 60 female students of the Gakushuin University in Tokyo. The Canadian group consisted of 30 female and 54 male students from the University of British Columbia. In general, the results indicated that Japanese Ss decode the nominal stimulus of TAT pictures in significantly different ways from Canadian Ss. Examination of these differences does not support the proposition that Japanese Ss decode the world in socially stereotyped

ways. For example, the Japanese proportions for the "suicide" code was greater on Card 3BM, but smaller on Card 17GF than for the Canadian sample. The Canadian proportion for the "spying" code was greater on Card 9GF but smaller on Card 5 when compared with those of the Japanese. On the other hand, some of the results do seem to reflect the strong cultural pressures within the Japanese culture upon perceptual codes. Without exception, the Japanese sample has larger proportions in the "sadness," "aggressive," and "disphoric" codes than does the Canadian group.

It would seem, then, that culture is a significant determinant of the functional stimulus in a thematic apperceptive technique. Although the data on differences in perceptual codes between subgroups within one culture are scant, Veroff's (1961) data imply that blue-collar workers and white-collar workers may have differential perceptions of the stimulus meaning of achievement oriented pictures. Although problems of interpretation are posed by his study because no data were collected on the specific decoding categories used by his two groups, the finding that n Achievement scores based on pictures of blue-collar workers, in comparison with pictures of white-collar workers, elicit more productive relationships to other variables for white-collar workers than for blue-collar workers suggests indirectly that the coding reactions of the two groups are probably different to the blue-collar pictures. In the same study, Veroff also showed college-educated women may interpret the functional stimulus of a picture of a career work setting differently from grade school–educated women. While the high school– and grade school–educated females appeared to interpret a career picture in terms of achievement, college-educated women viewed this picture as a cue for achievement-conflict. This general finding is in accordance with the notion that the functional stimulus of a career picture is not the same for the two groups.

In the light of the foregoing findings and considerations, any valid comparison of two different groups must ensure that the functional stimulus of thematic apperception pictures is the same for both groups. That is, the decoding systems to pictures must be the same for different groups, if the results are to be attributed to personality factors.

For some obscure reason, cross-cultural research and studies comparing groups within a society continue to be marred by the failure to provide methodological control over the functional stimulus. While it is invidious to single out any one research from a host of studies as an illustration of the confounding effects of the functional stimulus, a recent contribution (Bradburn, 1963) will be used to provide final clarification of the conceptual and methodological questions posed by our

second postulate. On the basis of previous research, Bradburn hypothesized that father dominance is associated with low achievement in men. Since interview data indicated that Turkish fathers are more dominant than American fathers, he predicted that Turkish junior executives, as compared with American junior executives, would have significantly lower achievement. Both groups told TAT stories to six pictures described in Atkinson (1958) and the data show that the American sample had significantly higher achievement scores. While it was Bradburn's conclusion that the data supported his hypothesis, such a motivational explanation of the results only follows if the investigator can rule out the possibility that the six pictures possess trans-cultural equivalence for achievement cues. It is entirely possible in this study that the average differences between American and Turkish males are due to the differential functional stimulus strengths that the six American-constructed pictures had for the two groups. That is, the obtained differences might be due to the fact that the nominal picture stimuli could be more readily decoded as reflecting achieving cues by American than Turkish subjects.

Summary. Research aimed at between-group comparisons must establish that projective stimuli evoke the same decoding systems for the groups. Unless the investigator controls for trans-group equivalence of stimulus meaning, his results must be viewed with reservations.

C. POSTULATE THREE: PROJECTION AS A FUNCTION OF PERCEIVED SIMILARITY

Attribution of tension-producing affects in awareness will increase as a function of increasing perceived similarity between the individual and the projective stimulus; attribution of tension-producing affects not in consciousness will increase as a function of decreasing perceived similarity between the individual and the projective stimulus.

Most recently, the empirical evidence and theorizing of a trio of investigators (Feshbach and Feshbach, 1963; Feshbach and Singer, 1957) have indicated that affective attribution is dependent not only upon the affective state of the perceiver but also upon the nature of the perceived stimulus. The following discussion takes as its frame of reference a recent theoretical article by Singer (1963).

In a pioneering and ingenious experiment on projection, Murray (1933) showed that frightened female children perceive pictures of adults as malicious. Murray called this type of affective attribution "complementary projection." The attribution to a stimulus figure or object of affect present in the perceiver may be called "supplementary projection." Unfortunately, Murray did not assess for the possibility of supplementary

projection effects in his experiment, nor did he attempt to determine how projection is critically dependent upon stimulus variables.

In the first of their studies, Feshbach and Singer (1957) used three groups, a control, a fear-expression group (who were encouraged to become conscious of their anxiety feelings), and a fear-suppression group (who were encouraged to suppress their affect). In support of their predictions, both fear groups projected fear onto a male graduate student appearing in a film, with the fear-suppression group ascribing more anxiety to the stimulus figure than the fear-expression group. The two most interesting features of these findings are that supplementary projection was demonstrated and that suppression of affect facilitates projection. In another follow-up study on the projection of anxiety, Singer and Feshbach (1962) showed that anxiety-aroused normals and psychotics attributed more anxiety to a series of human pictures than did control groups. More importantly, the specific prediction that psychotics would attribute more fear to a set of pictures labeled as "mentally ill" persons than to "never mentally ill" stimuli was supported.

Since these two experiments and Murray's left partially unanswered the question of what conditions generate supplementary or complementary projection, Feshbach and Feshbach (1963) hypothesized that the type of projection manifested will be contingent upon the relationship between the perceiver and the stimulus figure being perceived. Specifically, they predicted that supplementary projection will occur to stimulus figures viewed as similar to the perceiver, but that complementary projection will occur to stimulus figures perceived as dissimilar. In the study, young boys judged 16 pictorial stimuli, equally divided among men, women, boys, and girls, for the personality characteristics of maliciousness and fearfulness. In accordance with their predictions, the fear-aroused children, as compared with the controls, manifested complementary projection to the adult male stimuli and supplementary projection to young boy pictures. Such results strongly suggest that if the stimulus figure is perceived as similar to the perceiver, then direct attribution of affect is to be expected; on the other hand, dissimilar stimuli will instigate complementary projection.

The findings from the foregoing experiments indicate that anxiety affects will be projected in accord with the third postulate. Will the assumption apply to other affects? In an unpublished study discussed by Singer (1963), it was shown that college males and females who were made hostile attributed more hostility to college-age male stimuli than to a 12-year-old boy pictorial stimulus. In another study examining the interactions between the similarity of the perceiver and the projective stimulus, Feshbach, Singer, and Feshbach (1963) had college students

describe 26 stick figures and 28 TAT-like two-person interaction scenes. Both classes of stimuli conveyed the impression of hostility. In the TAT-type pictures, one person could be seen as frustrating another. In one series of the TAT pictures, five of the situations portrayed the receiver of the frustration as a college-age male, and in the other four situations he was a 12-year-old boy. In the other series of the TAT pictures, the ages of the stimulus figures were reversed, with the object of frustration and its source similar to those in the first series. While there was no difference between experimentally aroused hostile Ss and controls in attribution of hostile affect to the stick figures, both male and female angered Ss showed a heightened increment in the ascription of hostility to same-age college male stimuli on the TAT pictures. The experimentally aroused Ss showed no significant increment in hostility attribution to the younger stimulus individuals on the TAT pictures. Thus, the major findings of this study warrant the conclusion that hostile attribution only occurs to same-age figures and not to stick figures or to younger boys. The conclusion provides evidence in support of the proposition that supplementary projection is contingent upon the perceived similarity between the stimulus and the perceiver.

In another study in this series, Kiesler and Singer (1963) reported that college female Ss, who are made hostile but remain relatively unaware of their affects, attribute hostile feelings more to dissimilar persons than to similar individuals. This finding is consistent with the notion that affects not in awareness will be attributed to dissimilar stimulus figures.

In the most recent of this provocative series of investigations, Feshbach (1963) showed clearly that supplementary projection takes place for a socially acceptable and positive affect, namely, that of "happiness." Twenty-nine female Ss were made to feel happy by the joint interaction of a monetary reward and successful performance on a jigsaw puzzle designed to measure spatial ability. Twelve of the 29 Ss were assigned to an affective restraint group by instructing them not to reveal their feelings to other Ss. The experimental subjects, plus an additional 27 females in a control group, then evaluated for happiness 14 photographs of male and female pictures which had been previously scaled by a modified Thurstone technique on a continuum of happiness-sadness. At least one male and female photograph appeared at each point on the scale. The results of the ratings supported the hypothesis that a positive affect may be projected in that the affectively restrained Ss attributed more happiness to the happy-appearing female photographs than the controls. That supplementary projection depends upon similarity between stimulus figure and perceiver was also upheld, as shown by the

finding that there was more attribution of happiness to female faces than to male faces.

The results of all these studies are, in general, congruent with what would be expected in terms of the third postulate. It should be emphasized that this assumption should hold even when socially unacceptable affects are at issue. A particularly striking example of the independence of this assumption from the social acceptability of affects is provided by the research of Bramel (1962). In his experiment, Ss were made to feel that they were homosexually aroused by male photographs. This procedure caused the Ss to overestimate the magnitude of homosexuality toward individuals favorably evaluated as belonging to the same social category. While the latter type of individual may be assumed to be more similar than dissimilar to the perceiver, Bramel (1963), in a partial replication of the original experiment, attempted to manipulate more explicitly the variable of similarity and dissimilarity. In this experiment, Bramel employed the same technique as before to make the experimental Ss believe that they had homosexual motivation by having them observe their own deceptively high galvanic skin responses when viewing photographs of good-looking men in states of undress. Before watching the galvanometer dial, the experimental Ss had been led to believe that high galvanometer needle readings indicated homosexual arousal. For the control Ss, the needle remained at the low end of the galvanometer dial. After all Ss viewed and recorded their own need readings for 15 male photographs, they rated the degree of homosexual tendencies of an individual telling stories to four TAT pictures. The actual stories contained no overt references to homosexual or heterosexual behavior. Before listening to the tapes of the TAT stories, half of Ss (the "student group") were told that the story teller was a student; the other half (the "criminal group") were told that the story teller was a criminal who had just been arrested for the third time after a series of armed robberies. In addition, all Ss rated themselves on homosexuality and how they compared themselves in similarity to the story teller. The latter ratings were performed to assess the adequacy of the induction of the two independent variables—homosexual arousal and the variable of similarity-dissimilarity. Bramel found that the experimental Ss ascribed more homosexuality to themselves, reflecting the adequacy of the experimental arousal of homosexuality; the Ss assigned to the student group rated themselves as more similar to the story teller than did those in the criminal group, indicating the adequacy of the similarity-dissimilarity induction. In conformity with the third postulate about projective techniques, Ss in the student group attributed more homosexuality to the story teller than those in the criminal group. Thus, supplementary projection takes place to stimulus figures perceived as similar

to the perceiver, even though the attributed conscious trait is undesirable.

It is evident that the third postulate and its experimental evidence have profound implications for projective techniques. The fruitfulness of this assumption, over alternative ones, will now be considered with special reference to a sampling of related experimental and clinical problems in the field of projective techniques.

1. Hero Assumption

In the clinical use of thematic apperceptive techniques and doll play, it is sometimes assumed that the hero of a TAT story or the central identification figure in doll play reflects the characteristics of the subject and that the other figures in the story or doll play indicate how the subject perceives other individuals in his environment. Lindzey (1952) considers the hero assumption as a necessary feature of the clinical use of the TAT. On the other hand, Piotrowski (1952) has made the assumption that all figures in a story are equally revealing of the story teller's personality. The large-scale research endeavors of McClelland *et al.* (1953) on the achievement motive follow Piotrowski's assumption on this issue. Surprisingly, however, this important problem has generated very little direct research.

Lindzey and Kalnins (1958) undertook the only two studies on this problem, making the assumption that a distinction should be made between hero and nonhero figures in a TAT story. In the first of these studies, 30 undergraduate females were asked to evaluate the similarity between themselves and each character that they introduced into the six TAT stories they told. For the 180 stories, two independent raters agreed 90% in identifying the hero figure in each story. The main results of this study showed that the hero figures were more often perceived as similar to self than were nonhero figures. While Lindzey and Kalnins accepted this finding as offering modest support for the hero assumption, it must be pointed out that their study is not free from methodological criticism. In any valid test of the hero assumption *versus* alternative assumptions, it seems necessary to obtain independent evidence of the attributes of the subjects and then assess the extent to which these attributes are ascribed to heroes and nonhero figures. Without this kind of data, the Lindzey and Kalnins study is most difficult to interpret. If independent data on the story teller are obtained, then the hero assumption would predict that hero figures in TAT stories would be ascribed more of the story teller's attributes than would nonhero figures. Piotrowski's assumption would have to predict that these attributes be attributed equally to all figures. On the other hand, our third postulate would predict that attribution of traits would depend critically upon the perceived similarity of the story

teller to the stimulus figures in the TAT pictures and the degree of consciousness of the attributes being projected. This type of experiment has yet to be done.

In the second study, Lindzey and Kalnins (1958) exposed the experimental Ss to a complex frustration of social motives and compared their TAT protocols to those of unfrustrated controls. They hypothesized that the critical comparison between the hero assumption and Piotrowski's alternative assumption (i.e., that all figures are equally revealing) would be for the category of aggressive acts by nonhero figures in the stories. They argued that the hero assumption, in contrast to the aternative assumption, would predict no increase in aggression between nonhero figures. The data confirmed the prediction deduced from the hero assumption. However, our third postulate suggests that the results for the category of aggression by nonhero figures could be made to vary experimentally, provided the perceived similarity between these figures and the story teller were manipulated appropriately.

In light of the experimental findings that affects can be attributed to persons who are quite dissimilar to the perceiver, the hero assumption does not seem warranted in the interpretation of TAT stories. The inference from our postulate is that the story teller is quite likely to attribute affects to nonhero figures who are perceived as dissimilar to him. This observation should not be construed as implying that TAT protocol analysis should not attempt to distinguish between hero and nonhero figures. Quite the contrary. Attributes ascribed to nonheroes, provided they are dissimilar to the story teller, are likely to reflect unconscious states. Attributes attributed to heroes, provided they are similar to the story teller, are likely to indicate conscious states. In essence, clinical interpretation of any projective material must be based upon the meaning that stimuli have for patients.

2. *Assumptions with Respect to Physical Similarity between Stimulus Figures and Projecting Individual*

We turn now to the theoretical and practical problem of what type of stimulus figures an individual can most readily identify himself with, thereby facilitating maximum projection. Most of the research on this problem has been designed to test the premise that there is a monotonic relationship between amount of projection and increases in physical similarity between the story teller and the figures on apperceptive pictures. The reader will recall that Murray's (1943) original division of his pictures into sets for men, women, boys, and girls assumes that age and sex similarity of the TAT figures will facilitate identification and projective mechanisms. Since then, Murray's assumption has served as a useful model

for many other apperceptive test constructors who have designed special sets of pictorial stimuli. In this connection, Henry's (1947) set of line drawings for Hopi and Navaho children, Thompson's (1949) series of pictures with Negro figures, the Michigan Picture Test (Hartwell, Hutt, Andrew, and Walton, 1951) for grade school children, the set of pictorial cards depicting nuns developed by Lasago y Travieso and Martinez-Arango (1946) for the personality assessment of nuns, and Kagan's (1956) 13 pictures to test aggression in boys between the ages of 10 and 13, may be cited as representative of authors who have, by varying degrees, accepted the Murray model.

Although a considerable amount of energy has been devoted to an examination of the assumed relation between physical similarity and projection, the obtained data present a problem in interpretation. In order to measure amount of projection attributable to differences in physical similarity between the projector and the stimulus figures in the pictorial stimuli, experimenters have commonly employed as their measures of the dependent variable, indices of story length, word counts, number of different themes, transcendence scores, level of productivity, and compliance with test instructions. However, there is little in the way of theoretical or experimental justification for the assumption that word counts and the like are valid or relevant measures of projection. Just because one set of pictorial figures evokes a higher word count or transcendence index than another set is by no means a relevant proof that the subjects are projecting more of their attributes to one of these sets. Such data, at least to this author, are irrelevant to the problem at hand. In order to test adequately the relation between physical similarity and projection, it is necessary either to induce personality states experimentally or to obtain independent evidence that the Ss manifest certain personality characteristics. Then, it must be shown that one kind of stimulus material is a better reflector of these attributes than another type of material. As a consequence of the failure to obtain prior data on the attributes that individuals might project, the validity of the studies about to be reviewed is of questionable character. For organizational purposes, the studies will be grouped under the categories of identification with animals, color-of-skin identification, bodily similarity identification, and same-sex and opposite-sex identification. However, all studies deal directly or indirectly with the problem of the relation between the projector and the stimulus figures on pictorial stimuli (see also Murstein, 1961, 1963b).

a. Identification with Animals. Ever since Bellak designed the Children Apperception Test (CAT), experimenters have been unusually vigorous in testing a subsidiary premise upon which this test was built. Bellak made the assumption that young children would more readily

identify with animal figures than human and it is this assumption which investigators have tested extensively. For obscure reasons, the central assumption of this technique, namely, that stimuli depicting eating, sibling rivalry, primal scenes, toilet settings, sleeping, and aggression evoke the cardinal problems of children, has remained untested.

In an early study with children between the ages of 5 and 10 years, Bills (1950) found that rabbit pictures instigated significantly longer stories than TAT cards. Moreover, the psychological inferences which might be drawn from these two tests appeared to be quite different. In a subsequent study, Bills, Leiman, and Thomas (1950) showed that neither the TAT nor the rabbit pictures produced material which agreed with data derived from play therapy sessions. In these two studies, adequate control was not exercised over the background features of the two sets of pictures.

In the first systematic attempt to control for the stimulus values of the human and animal pictures, Biersdorf and Marcuse (1953) had an artist construct six animal pictures similar to those on the CAT plus another six comparable pictures with humans substituted for animals. They attempted to ensure that the two sets of pictures were identical in emotional expression and in most other respects. When the two tests were given to 30 first graders, no differences were found between the techniques on such indices as number of ideas, number of words, reaction time, number of figures introduced into the stories who were not in the pictures, number of figures referred to in the pictures, and total response time. Two main criticisms of this kind of research are worth reiterating. First, it has already been pointed out that measures such as word count in no way indicate whether or not the Ss are projecting attributes which they possess. On this score alone, this and similar experiments do not provide a valid test of Bellak's proposition. Second, the human pictures indicate far more clearly to the Ss the sex of the figures than do Bellak's animal pictures. If animal and human pictures are to be matched, ambiguity with respect to sex of the figures must be constant.

Two additional studies have been conducted with the Marcuse pictures. Mainord and Marcuse (1954) gave the two sets of pictures to emotionally disturbed boys and girls whose mean age was 7 years. Again, simple quantitative indices showed no differences between the two kinds of pictures, although five clinicians indicated that the protocols based on the human pictures were clinically more useful than those based on the animal series. Interestingly enough, however, Budoff (1955) found that less well-adjusted children were more productive on the CAT pictures than on another set of comparable human pictures. In Budoff's study, normal children produced more revealing material on the human

series. Furuya (1957) administered the Marcuse pictures to 72 Japanese children, with an age range of 6 to 12 years. On 20 tests of significance, the human pictures yielded better results than the animal pictures for four indices.

Light (1955) compared CAT cards with TAT cards, using 9- and 10-year-olds. While no significant differences in number of words occurred when the two sets of cards were compared, the CAT cards were more productive on a number of other criteria, such as expression of feelings, number of conflicts, and themes. Since this study failed to match the animal and human pictures on a number of critical stimulus dimensions, and contained no data on the personality attributes of the subjects which might be projected onto the respective figures, the findings are not pertinent to the general problem at issue. Using a group of intellectually superior children in the first, second, and third grade in a comparison between the CAT and a somewhat similar set of human pictures, Armstrong (1954) found that the human pictures elicited a significantly greater number of statements that went beyond picture description. There were no differences, however, between the two sets of pictures on word count.

In the most recent study on this proposition, Weisskopf-Joelson and Foster (1962) attempted to remove some of the deficiencies contained in the foregoing researches. They used 40 kindergarten children, with a mean age of 6 years, 2 months. While the CAT is designed for children between the ages of 3 and 10, Bellak and Adelman (1960) feel that the issue of animal *versus* human pictures should be tested on young children. Many of the other studies have used children close to the 10-year limit. In addition, this study avoided the use of a counterbalanced design in which all Ss receive each animal picture and its matching human picture. Counterbalanced designs may contaminate the results with strong differential carry-over effects from analogous pictures. In order to match four CAT pictures with similar human pictures, the investigators were forced to modify, probably in critical ways, the CAT pictures. For example, two important changes were the addition of clothes to the animals and the omission of the mouse in Card 3. While some children identify with the lion in Card 3, many take the mouse as the identification figure. While it would probably be inappropriate to design a picture with a human figure appearing out of a small hole in a baseboard, the omission of the mouse means that the investigators are no longer examining the CAT. It would have been better to use one of the remaining six cards. The results of this study showed that the mean transcendence index of the human pictures did not differ significantly from that of the animal pictures.

While most of these researches tend to show either no difference between animal and human pictures or that the human pictures are slightly more productive, the limitations of these studies prevent the acceptance of the proposition that children will identify as readily or more readily with human figures and will project more onto humans. As already stressed, none of these investigations had independent information on the attributes which may have been projected. Researchers have yet to show sufficient ingenuity in ensuring adequate control over background factors in designing their sets of pictures. Research has been largely limited to older children. Moreover, some studies have employed counterbalanced designs in which going from A to B may not be the same as going from B to A. In short, this writer adopts the strong position that no study has valid data. This conclusion should be contrasted with Murstein's (1963b) assertion that, after reviewing the same studies, he had "laid to rest" the assumption about identification with animals. Bellak's hypothesis is deserving of more inventive research and treatment than it has received up to this time.

Finally, it should be mentioned that no research has been conducted on the relative merits of animal and human stimuli in projective films. In contrast to still picture apperceptive techniques which commonly use human stimuli, filmed apperceptive techniques have primarily made use of puppets (Haworth, 1960; Lerner and Murphy, 1941) and animated cartoons (Gemelli, 1951). Apparently, the Test Filmique Thématique (Cohen-Séat and Rebéillard, 1955) is the only filmed technique that has exploited the possibilities of human stimuli. The research possibilities of filmed projective methods, contrasting the projective merits of various classes of stimuli figures, are almost limitless. Haworth (1960) should be consulted for a comprehensive review of films as a projective technique.

b. Identification with Color of Skin. The question of whether Negro Ss might more readily identify with Negro figures on an apperceptive test was raised by Thompson (1949). He hypothesized that sparsity of dynamically relevant material in the protocol data of Negroes might be due to the white figures on the TAT. Accordingly, he substituted Negro figures for the white ones on the TAT and gave the new set of pictures, along with the analogous 10 pictures from the original TAT, to a group of 26 Southern Negro college students. The students produced significantly longer stories on the Negro series than on the traditional TAT, suggesting that projection is facilitated by color-of-skin similarity between the story teller and the figures in the pictorial stimuli.

However, Thompson's modification of the original TAT for Negroes was quickly challenged. Riess, Schwartz, and Cottingham (1950) found no differences in story length between the Negro and white version of the

TAT for 30 Northern Negro and 30 Northern white college students. Analyzing the same data, Schwartz, Riess, and Cottingham (1951) showed that Negro students produce the same number of ideas to both sets of cards. Korchin, Mitchell, and Meltzoff (1950) contributed another study on the two versions of the test. Again, story length did not differ between the two versions for 80 Northern Negroes and 80 Northern whites. Similarly, Light (1955) did not find any differences in story length between the TAT and the Thompson modification for 26 students. In a more appropriate experimental design that avoids the difficulty of the counterbalanced designs used in the other studies, Cook (1953) tested one half of his Negro and white Ss with the Murray TAT and the remaining Ss with the Thompson modification. In agreement with other studies, Cook found no significant differences due to version of the test.

Aside from Thompson's initial study, the main conclusion of these studies appears to be that physical similarity of skin color does not increase projection. However, these investigations suffer from the same limitation as those dealing with the animal *versus* human identification problem. In point of fact, there is no evidence that the Ss are attributing any attributes to either version of the TAT. Moreover, the assumption that any simple manipulation of skin color will elicit perceptions of similarity and identification is tenuous, to say the least.

c. *Identification with Bodily Similarity.* A few investigators have attempted to test the formulation that bodily similarity between story teller and the central figure on TAT cards will increase the amount of projection. The measures of projection are usually word counts, transcendence indices, or level of productivity. The principal criticisms of these studies concern the correctness of the measures of projection and the crude manner of manipulating bodily similarity. As we have already indicated, the only way in which an experimenter can be sure that a subject is projecting is to have prior knowledge of the subject's attributes. The similarity variable is something that resides within the subject. That similarity is a psychologically mediated variable and not a physical variable is illustrated by the finding (Clark and Clark, 1947; Goodman, 1946) that Negro and white preschoolers identify more strongly with the white doll than the Negro doll when asked, "Which one looks like you?" Thus, the kind of person an individual perceives himself similar to may not correspond with the experimenter's preconceived notions. For example, when an experimenter draws a TAT figure as obese, there is no guarantee that an obese person will perceive himself or his body image as similar to the obese figure in a TAT picture. This kind of research requires phenomenal judgments from the subjects as to the kinds of individuals they perceive themselves similar to in specifically defined role situations.

Armed with such judgments, the researcher could then manipulate the similarity variable in role settings represented on TAT pictures. With these reservations in mind, the three studies which have dealt with the body similarity problem will be quickly reviewed.

Weisskopf and Dunlevy (1952) obtained no significant interaction for transcendence indices between type of subject variable (obsese, crippled, and normal) and type of central figure variable (obese, crippled, and normal). In a second study, Weisskopf-Joelson and Money (1953) investigated the effects of facial similarity between subject and central figure in the TAT on word counts and transcendence indices. In the neutral set of nine pictures, the facial features bore no resemblance to the subjects. In the experimental set, the facial features of the central figures were the photographic reproductions of the Ss' faces. The results showed that facial similarity between the S and the central figure did not produce a significant change either in word counts or transcendence indices. Greenbaum, Qualtere, Carruth, and Cruickshank (1953), in an investigation comparing TAT pictures and modified versions showing stimulus figures as physically handicapped, found that the unmodified TAT cards yielded a greater level of depth than the modified set.

When considered against the bench marks established for this kind of study, it must be concluded that adequate experimentation has yet to appear concerning the relation between perceived bodily similarity of the story teller and the central figure.

d. Identification with Same-Sex and Opposite-Sex Figure. At the level of theory, it is often assumed that Ss more readily identify with stimulus figures of their own sex. Primarily on the basis of developmental identification with the same-sex parent and the sex typing of behavior in socialization practices, this assumption appears tenable. Thus, it seems reasonable to hypothesize that an individual will project tension-producing affects of which he is aware onto the same sex stimulus. However, tension-producing affects of which he is unaware should be projected onto opposite-sex figures. Does this twofold supposition conform to the experimental findings?

In an early use of a filmed projective technique, McIntyre (1954) concluded that a viewer does not increase projection to the like-sex film protagonist. A similar finding was obtained by Silverstein (1959), who administered two parallel sets of male and female apperceptive pictures and a shortened version of the Edwards Personal Preference Schedule to 40 male and 40 female college students. The thematic material and the Edwards schedule were scored for affiliation, aggression, autonomy, deference, exhibition, nurturance, sex, and succorance. The correlations between the Personal Preference Schedule scores and those from the the-

matic apperceptive pictures for the same-sex and opposite-sex figures were, in general, not significantly different from zero. Two major considerations vitiate any firm conclusions from these studies. Both used self-questionnaire data as the basis for information about the attributes of the subjects. Furthermore, the Ss in both studies did not seem to have identified with either sex figure in the projective material.

Moore and Schwartz (1963) have recently reported a comparison between two modified versions of the Rosenzweig P-F Study in which the sex of the frustrated figure was manipulated. Each version of the test consisted of 24 cartoons, with Form I so designed that a male figure was frustrated in the first 12 cartoons and a female figure was frustrated in the last 12 cartoons. In Form II, the female figures were frustrated in the first 12, and males in the next 12. Forty-eight males and 52 females received Form I and 52 males and 51 females responded to Form II. In contrast to the TAT, the P-F Study requires the S to identify with the frustrated person and to provide an appropriate response for him. The seven Rozenzweig scoring categories, when applied to the frustrated figures of the same sex and to those of the opposite sex, revealed no significant findings. The authors concluded that this contradicts the assumption of a relationship between attributes of the subject and pictorial stimuli. However, since the experimenters did not have any independent knowledge about the personality attributes of their Ss, their results are not germane to the assumption under consideration.

While the preceding studies have not yielded valid evidence on same-sex and opposite-sex identification, there are others which leave little doubt that sex of the stimulus figure is related to identification and projection.

Maccoby and Wilson (1957) carried out a study in which they tested the specific prediction that seventh grade pupils will identify with the like-sex character in a Hollywood feature ("The Tomboy"). This movie was thought to provide comparable opportunities in the choice of either a boy or a girl as the identificand. An identification index showed that the male viewers identified with the boy character and female viewers with the female character. Comparable results in another study (Albert, 1957) strengthen the supposition that Ss identify most strongly with same-sex stimulus figures.

However, it is possible for the individual to choose an identificand of dissimilar sex. While sex identity undoubtedly helps to mediate perceived similarity between the subject and stimulus figures, similarity cannot be measured by sex alone. The general role setting in which the figures are embedded, the sex typing of the behavior portrayed by the stimulus figures, the perceived consonance between the motives and ac-

tions of the stimulus figures and the subject, and the general adequacy of appropriate sexual identification within the subject are four important factors which we assume to mediate perceived similarity. These four variables must be taken into account before any adequate predictions can be made about identification with same-sex and opposite-sex figures. For example, if a male stimulus figure is engaged in sex-typed female activity, then a male subject may identify with an opposite sex figure, particularly if she is engaged in masculine activity. We also suspect that the identificand may be dissimilar in sex to the subject when tension-producing affects are not in awareness or when the subject's strongest developmental identification is to a member of the opposite sex.

It is these considerations that may help to clarify two of the most challenging and puzzling problems about achievement motivation in women. As a general background to research on the achievement motive, it should be mentioned that males have served as Ss in most of the studies that have demonstrated lawful relationships between achievement motive scores and performance variables. Studies on females have been few in number, largely because the preliminary research with females yielded equivocal data. As Lesser, Krawitz, and Packard (1963) have indicated, a comprehensive theory of the achievement motive should be able to incorporate the dynamics of achievement for both sexes. A theory of motivation, based on the findings of one sex, would be limited to approximately half the population.

In 1953, the two enigmas about the findings on the achievement motive for females were (a) why females did not increase in achievement under achievement-arousing operations, whereas males did; and (b) why women produced more achievement to pictures of males than to pictures of females. A decade ago, McClelland et al. (1953) reviewed three studies which showed that females did not show the expected increase in achievement as a consequence of achievement-involving instructions. Studies by Veroff (1950) and Veroff, Wilcox, and Atkinson (1953) indicated that female high school Ss and female college Ss manifest more achievement to male picture cues than to female ones. This finding occurred under neutral and achievement-oriented instructions, although neither study produced a significant increase in achievement motivation under arousal conditions. However, an examination of three recent studies with female Ss which have shown an increase in achievement with achievement-oriented instructions suggests an important interaction effect between the achievement motive and picture cues for females.

Angelini (1955) (reported on by Lesser et al., 1963) reported that competitive Brazilian college women showed a significant increase in achievement motive scores after achievement-arousal instructions which

stressed appeals to leadership ability and intelligence. More significantly, perhaps, the Brazilian college women obtained higher achievement scores to pictures of males than did Brazilian college men to pictures of males. It is important to note that Angelini used only male card stimuli.

Alper (1957) showed that the sex of the central figure in a TAT card has an important influence on achievement for females. In her study, three groups of college girls were used, a group composed of 28 leaders who were tested under neutral instructions, a second group of 24 nonleaders who were tested under ego-oriented instructions, and a third sample of 24 nonleaders who were tested under relaxed conditions. Card 1 (boy and violin) of the TAT and Card 11G (young girl in empty class-room) from the Michigan Picture Test were used in the data analysis. Independent of pictures, Alper found that the nonleader ego-oriented group had significantly higher achievement scores than the relaxed non-leader group. The results for the male and female pictures showed (a) the leader group produced more achievement to the male picture than to the female picture, but no differences between the two pictures occurred for either the ego-oriented nonleaders or the nonleader relaxed group; (b) the male picture produced a significantly higher achievement score for the leader group than for the nonleader relaxed group, as it also did for the nonleader ego-oriented group in comparison with the nonleader relaxed group; (c) the female picture produced a significantly lower achievement score for the leader group than for the nonleader group, as it also did for the ego-oriented nonleaders in comparison with the nonleader relaxed group; and (d) the female picture produced signifi-cantly more achievement for the relaxed nonleaders than for the leaders. This complicated pattern of results indicates that a male picture is likely to arouse achievement more than a female for either female leaders or women who have been given achievement-arousing instructions. Since Alper did not have an achievement-aroused leader group for a direct comparison with a nonaroused leader group, this study merits replication with the addition of an aroused leader group.

The Lesser et al. (1963) study explored the interesting hypothesis that the failure of past investigators to arouse the achievement motive experimentally in females might be due to the fact that the women Ss who have been used were not overly concerned with achieving through means of intellectual skills. To test this hypothesis, they used 80 juniors and seniors from a high school for intellectually gifted girls. The ad-mission to this school is competitive, and, once selected, the student is placed in a highly competitive environment which places stress on intel-lectual and scholastic achievement. The Ss were divided into two matched groups of 40 achievers and 40 underachievers. All Ss were exposed to neutral and to achievement-oriented instructions. Under each of the con-

ditions, Ss wrote TAT stories to three pictures in which a male was the central figure and to three pictures in which a female was the central figure. While the main hypothesis of the study that the achievement instructions would yield an increase in achievement motivation was not confirmed, a significant interaction among groups, experimental treatments, and pictures was found. The interaction showed that "achieving girls did display the expected increase in achievement motivation scores under achievement orienting conditions, but only when responding to stimuli depicting females; underachieving girls also displayed the expected increase in achievement motivation scores under achievement orienting conditions, but only when responding to stimuli depicting males" (Lesser et al., 1963, p. 63). The authors applied a social role interpretation to their results. They made the assumption that achieving females conceive of intellectual achievement as an important part of their role, whereas underachieving girls perceive that the pursuit of intellectual achievement is part of the male role. In addition, they made the assumption that the attribution of the aroused motivation of achievement in the underachievers would be directed to male figures who are perceived as the proper agents for achievement expressions. Thus, achieving girls will attribute achievement to female figures and underachieving girls will attribute achievement to males. This social role explanation of the obtained results is most plausible and could be checked by obtaining role expectations that achieving and underachieving girls have for males and females. On the other hand, our third postulate would propose that achievement strivings are less in awareness for the underachievers than for the achievers. If this were the case, then it could be predicted that underachieving girls would project more of their achievement strivings onto male figures than female figures.

Summary. There is strong experimental evidence showing that individuals project their affects in accordance with the third postulate. The application of this postulate to the hero assumption and the assumption with respect to physical similarity between stimulus figures and the projecting individual was examined in great detail. Unfortunately, most of the studies on these assumptions suffer from methodological weaknesses. However, the evidence does indicate that type of stimulus figure on the TAT plays a critical role in controlling type of attribution.

D. POSTULATE FOUR: PERCEPTUAL DYSFUNCTIONING AS A FUNCTION OF PROTUSIVE STIMULI

Acutely protusive stimuli will disrupt perceptual and cognitive organizing functions in individuals with weak decentering ability.
Contemporary research on the development of perception and cog-

nition (e.g., Piaget, 1950) shows that percepts and cognitions are initially under the control of the immediate stimulus environment. At birth and for sometime thereafter, objects possess no symbolic meaning, and consequently are perceived solely in terms of their color and shape. Early perceptual responsivity is immediate and spontaneously directed toward the environment. With increasing perceptual maturity, however, the impact of the immediate stimulus upon perceptions and cognitions becomes less and less. Proximal stimulation becomes subordinated to decoding systems and to abstract schemata. When a child has reached the stage of concrete and formal operations, he is no longer under the sway of the salient features of stimuli. He can decenter or refocus on various aspects of the immediate external and internal environment by highly flexible and balanced modes of perceiving.

From these notions, it follows that a large number of decoding systems and good decentering ability are the two most important factors that permit easy accommodation to protusive stimuli. While what constitutes a protusive stimulus for any individual can only be discovered by experimental techniques, it is suggested that any stimulus that cannot be readily assimilated into decoding systems is likely to function as a protusive cue. Kreezer's (1958) threshold technique for measuring the attention demand value of a stimulus is one possible operational method for determining protusive stimuli. Bright colors and affectively charged or threatening figures are two examples of stimuli which may be protusive for many individuals. For Ss with good decentering ability, such stimuli would, however, be incorporated into an existing decoding system. On the other hand, individuals with minimal decentering ability may find such stimuli hard to incorporate into a decoding system. In such a case, protusive stimuli are assumed to raise the individual's arousal level to such a point that it disrupts smooth perceptual functioning (see Section II, H and I, for a discussion of arousal level effects). It is further assumed that arousal level is lowered by resort to an early mode of perceptual responding, that is, by response immediately and directly toward the salient features of protusive stimuli.

This formulation implies that the effects of protusive stimuli are not likely to be shown on projective techniques unless individual differences in decentering ability are taken into account. If the decentering variable is overlooked in an experimental design, then the effects of protusive stimuli will not be manifest.

The bridge between these theoretical notions and projective methods will be limited to the possible effects of bright color as a generalized protusive stimulus. As applied to the Rorschach, it is predicted that protusive colors may cause perceptual dysfunctioning in individuals with

weak decentering ability. The perceptual dysfunctioning is probably best assessed by combining the C and CF Rorschach determinants. Several factor analyses of the determinants have shown that the Rorschach does measure a lack of perceptual control factor. In both psychiatric and undergraduate students, Wittenborn (1950a, b) isolated a factor that involves a spontaneous perceptual approach to inkblots. The high loadings on C and CF indicate that spatial features of the inkblots are being minimized. The clustering of C and CF on a lack of perceptual control factor was also confirmed by Williams and Lawrence (1953). Inasmuch as the FC determinant did not participate in the loadings on this factor, it would seem appropriate to treat the personality implications of the FC response as different from CF or C responses.

Unlike many other hypotheses about color as a protusive variable, the present formulation is not rooted in the assumption that color evokes affective responses. In essential agreement with Schachtel (1943), Rickers-Ovsiankina (1943), and Shapiro (1960), the present hypothesis asserts that primitive color percepts represent a more immediate form of perception and is indicative of perceptual dysfunction. In basic agreement with Shapiro (1960), the present formulation remains theoretically neutral about the relation between color and affect.

1. Evidence That Primitive Color Percepts Represent an Immediate Perceptual Mode and Are Reflective of Perceptual Dysfunctioning

While factor analyses of the Rorschach determinants indicate that protusive chromatic stimuli may produce a lack of perceptual control factor, is there more direct evidence to substantiate the hypothesis that individuals scoring high on such a factor manifest an immediate and spontaneous approach to the environment? A few studies (Mann, 1956; Singer and Spohn, 1954; Clark, 1948) have provided data that support the notion of a linkage between color responses and a spontaneous mode of perceiving. Research (Bills, 1954; Costello, 1958; Kohler and Steil, 1953; Wittenborn and Holzberg, 1951) has shown that depressive individuals make little use of color, indicating a lack of spontaneous and immediate reactions to protusive stimuli. In their factor analysis of Rorschach determinants and MMPI scales, Williams and Lawrence (1954) also obtained a significant negative correlation between the Depression scale and CF and C determinants. Superficially there was a finding in this study which would appear to contradict the present theoretical expectations, if ego strength is viewed as a reflector of a perceptual capacity to synthesize and inhibit excessive stimuli. This finding was the correlation of .50 between CF scores and Barron's (1953) Ego-Strength (Es) scale. However, the construct validity of the Es scale

is open to question. For example, Gottesman (1959) found significantly higher Es scores for delinquents over normals. Barron (1956) likewise found that Es scores correlated with a cluster variable indicative of aggression. Moreover, in his original study Barron (1953) found that Es scores were positively correlated with the variables of outgoingness and spontaneity and negatively with introspection. Such data, if anything, indicate that the positive correlations between Es and CF support the present theoretical formulation.

2. Experimental Manipulation of Color on the Rorschach

The finding of Brosin and Fromm (1940) that color-blind neurotic individuals manifest "color shock" was the first report to raise the whole question as to whether color influences response performance. Our hypothesis suggests that color should have some effects. Since the early experimental endeavors of Wallen (1948) and Lazarus (1948, 1949), many investigators have studied color effects by comparing chromatic and achromatic versions of the Rorschach inkblots.

Since numerous other researchers (Allen, 1951; Allen, Manne and Stiff, 1951, 1952; Allen, Stiff and Rosenzweig, 1953; Barnett, 1950; Buker and Williams, 1951; Dubrovner, Von Lackum and Jost, 1950; Perlman, 1951; Sappenfield and Buker, 1949; Swartz, 1953) have followed the general methodological approach of Lazarus (1949), his test-retest experimental design merits critical evaluation. It involved (a) the administration of the standard Rorschach inkblots and their achromatic matches in a counterbalanced order to two groups of Ss, and (b) a statistical comparison of the standard chromatic series and the achromatic version on various scores that are thought to be symptomatic of color disturbance. Many critics have raised cogent criticisms against the test-retest design, of which the following seem particularly crucial: (a) the assumption that an AB order of presentation involves the same memory and set effects as a BA order of presentation is highly questionable; and (b) since color may either inhibit or excite perceptual reactions, depending upon a critical interaction between kind of color and personality structure, the analysis of data based upon only one group will lead to a general cancellation effect. While recent investigators have usually avoided the counterbalanced design, only a few have attempted to take into account the possible interaction effects among kind of color, receptivity to color, and personality structure. It is for this reason that most of the research studies manipulating color on the Rorschach experimentally are not pertinent to the personality and perceptual hypotheses about the protusive role of color. In view of this deficiency, no attempt will be made to discuss most of these studies in any great detail. However, Baughman

(1958) should be consulted for the details of the studies conducted before 1958.

  a. *Studies with Normal Subjects.* With the preceding reservation in mind, a brief review will be provided of the studies that have investigated the influence of color with normal *Ss*. Using 100 high school seniors, Lazarus (1949) found few differences between the achromatic series and the chromatic series. He did, however, find a higher *F%* and number of space responses on the noncolor cards. There was also a reduction in the number of *D* responses to the achromatic version. Since this experiment, several similar experiments (Allen, 1951; Allen, Manne and Stiff, 1951, 1952; Barnett, 1950; Canter, 1951; Dubrovner, Von Lackum and Jost, 1950; Grayson, 1956; Meyer, 1951; Sappenfield and Buker, 1949) with normals have shown few or no differences in response performance between chromatic and achromatic versions of the Rorschach.

  However, two recent studies indicate that color does have an influence on Rorschach performance. Exner (1959) investigated the role of chromatic and achromatic color on Card I of the Rorschach by constructing three different chromatic versions of this card by the monochrome-dye coupling method. The *Ss*, male and female students, were divided into four matched groups of one hundred cases each. The *Ss* in each group were administered a different version of Plate I. The chromatic versions of this card were blue, green, and brown and the achromatic version was the standard Card I. The data showed that the blue card yielded the greatest number of responses, the shortest reaction time, and significantly more wholes, small details, human movement, whole human figures, object content, "bell" content to Beck's *Dd24*, snow content, and animal face content than the other version of Card I. While it is a little disappointing that Exner did not include a red version of Card I, most of his data on the blue card is in keeping with the finding that blue is relaxing (Wexner, 1954), and produces less cortical and autonomic arousal than red (Gerard, 1958). The green card produced significantly fewer form determinants than any other version of Card I. These findings strongly support the conclusion that different colors have dissimilar effects. This conclusion, as Exner indicates, supports Crumpton's (1956) finding that a color perception on one Rorschach card may not indicate the same interpretation as a similar response on a different plate, and reaffirms the implications of the first postulate.

  In the most comprehensive analysis of the stimulus structure of the Rorschach to date, Baughman (1959) prepared six different modifications of the standard Rorschach inkblots. His findings on color-achromatic comparisons will be limited to a discussion of plates II, III, VIII, IX, and X. On Plate II, the chromatic series yielded fewer movement and

human percepts but more blood responses to the $D2$ area than any of the other card versions. On Plate III, color increased anatomical responses to the $D3$ area and lowered the number of responses to $D2$. Content analysis of Plate VIII showed that color depressed bear concepts, and catlike and landscape responses to the $D1$ area, and facilitated the butterfly, flower, and ice cream percepts to the $D2$ area. The chromatic series of Plate IX produced few minor details, but many whole percepts. Color also depressed the human percept to the $D4$ area. The colored version of Plate X produced an increased sensitivity to seed percepts to the $D3$ area, caterpillar or worm responses to the $D4$ area, rabbit head percepts to the $D5$ area, insect percepts to the $D8$ area, and flower percepts to the $D15$ area. While only a sampling of Baughman's data has been presented, two primary conclusions may be drawn: Color does influence Rorschach performance, especially the content categories; and color may either facilitate or inhibit responses.

The question now becomes one of accounting for individual differences in response to color. Conceptually, as already indicated, the studies of normals do not bear directly on the hypotheses relating color perception to personality variables. A design is required that will permit testing for the triple interaction effect among types of intrusive and non-compelling colors, receptivity to color, and personality structure.

b. Studies with Neuropsychiatric Subjects. Some researchers (Allen et al., 1953; Baughman, 1954; Buker and Williams, 1951; Canter, 1958; Crumpton, 1956; Sterling, 1950; York, 1951) have attempted to escape the limitations of a normal sample by using psychiatrically disturbed Ss. However, studies with a single group of patients are no closer to an appropriate design than studies employing a single group of normals. Both designs overlook the requirement that interaction effects should be built into the experimental design.

With this qualification in mind, three studies (Allen *et al.*, 1953; Baughman, 1954; Buker and Williams, 1951) have shown that the removal of color does not seriously influence response performance on the Rorschach. Canter (1958) showed that an unshaded bright color version of Plates VIII, IX, and X produced an inhibition of color-dominant responses. Accuracy of form perception was impaired in the York (1951) study. With a sample of 10 psychotics, 10 neurotics, and 10 organics, Crumpton (1956) showed that color has no influence on color shock signs, productivity, reaction time, 8-9-10%, rejections, minus form quality of responses, $P\%$, $A\%$, avoidance of red area on Plates II and III, and irregularity of succession. However, color did have effects on content categories, supporting the findings of Baughman (1959) and Exner (1959) with normals. Whatever advantage these studies might have gained by

the use of disturbed Ss was eliminated by failure to separate the patients in terms of personality variables hypothesized to be related to color perception. Thus, the findings from these studies cannot be related to any of the hypotheses about color on the Rorschach.

c. *Studies with Interaction Effects.* Five studies have contrasted normals and disturbed Ss on chromatic and achromatic versions of the Rorschach. Three (Brody, 1953; Perlman, 1951; York, 1951) of these studies have compared neurotics and normals on the two versions of the Rorschach and found that the effects of color are minimal or absent. While these studies approach the ideal design, they overlook the fact that the "crude" classification of normal or neurotic masks the individual differences within each of these groups which are likely to produce significant effects. It is well known that individual differences in autonomic and perceptual reactivity in neurotic and normal samples are likely to be large. Accordingly, the differences between each of the groups are likely to be canceled out, which ensures that they will not differ on most dependent variables. Rather than employing heterogeneous samples, one should use groups which are high and low on the specific personality variables assumed to be related to color perception, such as affect, inadequacy in perceptual decentering, and the like.

In a study that approximates the ideal design, Holtzman, Iscoe, and Calvin (1954) tested for relationships of type of color, anxiety, stimulus structure, presentation order, and examiner. Unfortunately, the high- and low-anxiety Ss were chosen on the basis of the Taylor Manifest Anxiety Scale (1953) and the Winne Scale of Neuroticism (1951). This means that the contrasting groups of college women were composed of those who admitted to neurotic symptoms and those who did not. In anticipation that type of color may produce an effect, the results of Plates II and III were analyzed together because their chromatic versions are both red and black. Similarly, Plates VIII and IX were grouped together because their standard versions involve multicolored pastel shades. A balanced experimental design was employed, in which each S received two chromatic and two achromatic inkblots in a predetermined order. Contrary to the theoretical expectations of some Rorschach workers, the findings did not support the assumed linkage between color and anxiety. Differences between the chromatic and achromatic versions were minimal or absent. Forsyth (1959) also found no significant differences when high anxious and low anxious male undergraduates were tested for differences on Elizur's content anxiety score. Anxiety was defined by the Welsh (1956) manifest anxiety scale and the comparisons were made between a chromatic, an achromatic, and a silhouette Rorschach series. A significant interaction effect showed that a moderately

anxious group produced a low Elizur anxiety score on the colored cards and a high Elizur anxiety score on the achromatic and silhouette Rorschach series. While the Holtzman *et al.* and Forsyth studies are neatly conceived, their findings are only suggestive and do not provide critical evidence on the possible interaction effect among color type, color receptivity, decentering activity, and personality structure.

d. *Studies with Segments of the Rorschach.* The effects of color on Rorschach performance deserves one final methodological comment. In the studies reviewed to this point, the complete Rorschach inkblot has been used. There have been, however, a few studies (Siipola, 1950; Siipola, Kuhns and Taylor, 1950; Lazarus and Oldfield, 1955) in which segments of the usual details have been used. While results from these studies have indicated at least a modest effect of color on response performance, the results of such studies may not apply to intact inkblots.

3. *Experimental Manipulation of Color on the TAT*

Color has been shown to have effects on thematic apperceptive techniques in four of five studies. Brackbill (1951) found that neurotics, unlike control Ss, told more depressive stories to chromatic TAT pictures. Thompson and Bachrach (1951) found that both whites and Negroes gave significantly higher words on the chromatic version of the TAT. Lubin (1955) also found significantly higher word counts on a chromatic TAT for mentally retarded Ss, as did Lubin and Wilson (1956) for handicapped children. On the other hand, color did not influence the transcendence index in the Weisskopf–Joelson and Foster (1962) experiment when color and its absence were varied on the CAT. In general, these studies only demonstrate that color facilitates word productivity. In no way do they provide crucial information on the interaction effects among types of color, receptivity to color, perceptual decentering, and personality structure.

Summary. Only indirect evidence supports the validity of the fourth postulate, namely, the finding that color perception represents an immediate approach to the environment. While recent studies have shown that color influences response performance on the Rorschach and TAT, adequate experimental research remains to be done on the hypothesis that protusive stimuli disrupt perceptual activity in Ss with weak decentering ability.

E. POSTULATE FIVE: ANXIETY AS A FUNCTION OF CHIAROSCURO STIMULI

Diffuse chiaroscuro stimuli evoke vague homogeneous perceptions which tend to be anxiety arousing.

Working with Binder's (1933) interpretations of chiaroscuro responses

on the Rorschach, Bohm (1960) and others have indicated that the two conditions necessary for the occurrence of a homogeneous sensory perception are similarity between stimulus elements and indistinct delineation. The latter condition is usually met by chiaroscuro stimuli. However, both conditions are met by exposing Ss to a homogeneous field (Ganzfeld). In order to achieve a Ganzfeld, similar to that described by Hochberg, Triebel, and Seaman (1951) and Kenny and Chappell (1963), Ss are required to wear eyecaps which are cut from table-tennis balls to fit the shape of the eye socket. A red light is projected onto the balls. Within a relatively short period of time, the initial red perception changes into an unpatterned gray visual field. It appears to the S that he is being stimulated by diffuse chiaroscuro stimuli and he may describe his perceptions as being foglike, cloudlike, or simply graylike. In short, the S has vague, homogeneous perceptions.

In terms of the present postulate, the induction of a Ganzfeld should be anxiety arousing. A recent experiment (Kenny and Chappell, 1963) has provided empirical support for this notion. Experimental Ss were subjected to a Ganzfeld for 20 minutes, and while still under this condition gave stories to six TAT card descriptions. A comparable control also told stories to the TAT card descriptions. The success of the Ganzfeld as an arouser of anxiety was assessed both by an anxiety questionnaire completed by the Ss after telling their stories and by scoring the TAT stories for anxiety effects. The anxiety questionnaire results showed the experimental Ss to be more anxious than the controls, and the TAT findings offered substantial support for the hypothesis that a continuous gray field is anxiety arousing. Six out of the 10 TAT anxiety indices showed significant differences in the predicted direction.

While this experiment offers evidence in support of the hypothesis that diffuse chiaroscuro stimuli evoke vague homogeneous perceptions which tend to be anxiety arousing, it only offers indirect support for any anxiety interpretation of chiaroscuro material on the Rorschach.

However, there is other evidence congruent with such an interpretation. In an analysis of the connotative meaning of the Rorschach, Rosen (1960) had his Ss make semantic judgments on the Rorschach inkblots and nine common Rorschach responses. The Ss rated the response "cloudiness" as bad, sad, hazy, cold, and dull. As Rosen indicates, such undesirable connotations are close to a description of generalized anxiety.

Ultimately, the validity of this postulate must be tested either by inducing anxiety or by using psychiatrically defined cases of anxiety and comparing their responses with those of matched controls. In terms of the notion of homogeneous perception, it would be desirable to limit the scoring of the shading material to Binder's *ChF* and pure *Ch*

responses. With such a stipulation, it becomes extremely difficult to assess the validity of the fifth postulate through an examination of the research relating anxiety effects to the Rorschach, as investigators have used diverse scoring criteria for shading. If citation is limited to studies that come close to Binder's two scoring categories, the studies of Levitt (1957), Cox and Sarason (1954), and Levitt and Grosz (1960) support the proposition that shading responses are correlated with anxiety. Such findings buttress the present postulates, especially when it is noted that the three studies used different kinds of Ss and different external criteria of anxiety.

Summary. There is indirect evidence to support the notion that diffuse chiaroscuro stimuli instigate homogeneous perceptions which tend to be anxiety arousing. However, a great deal more experimental work is required on the present postulate before its validity status becomes known.

F. POSTULATE SIX: IDIOSYNCRATIC RESPONSES

If one rules out differential cultural experiences, response deviations (idiosyncratic responses) from the perceptual demands of the projective stimulus are indicative of significant personality characteristics.

From a research and clinical point of view, two interlocking questions emerge from this general proposition:

1. Is reality-oriented behavior indicated by conformity to the demands of stimulus pull?

2. Are idiosyncratic responses more significant than usual responses?

To both questions, the answer appears to be affirmative. The first implies the assertion that the larger the response discrepancy from the stimulus pull of the projective material, the greater will be the impairment in the individual's ability to integrate perceptual and cognitive processes. Given the question on the Rorschach, "What might this be?," the individual has to attend to the stimulus, and, to the extent that the response is not stimulus bound, the inference may be drawn that perceptual and associational functioning are not being integrated. Putting this assertion the other way round, it is assumed that if the S is overresponding to his own associations and cognitions, then he is less likely to be attending to the reality demands of the stimulus situation. Lest the logic of this assertion be misunderstood, it should be stressed that it is not implied that compulsive adherence to the stimulus environment reflects ideal psychological health. The shape of the curve relating response adherence to stimulus demands and psychological health is undoubtedly nonmonotonic and close to an inverted U.

Many factors may reduce a subject's ability to integrate perceptual

and cognitive processes. Clinically speaking, the three basic factors that probably produce idiosyncratic response are minimal decentering ability, neurophysiological dysfunctioning, and high levels of arousal in the presence of ambiguous cues. Strong motives per se should not impair an individual's respect for reality, unless some variable like cortical dysfunctioning is also present.

Before examining the research literature on this hypothesis, one methodological point is worth raising. In any rigorous test of the postulate, it would be most important to select stimuli that have highly predictable responses, that is, stimuli with few associations. If a stimulus evokes many different responses, the experimenter will have problems in establishing criteria for idiosyncratic responses. A control over the number of associations is also important in research which seeks to demonstrate that particular kinds of stimuli produce idiosyncratic responses in Ss with special kinds of conflict. For example, it has been commonly assumed that schizophrenics, relative to normals, produce a large number of idiosyncratic responses to emotional words in a word association test. Two studies (Deering, 1963; Jones, 1957) have found that schizophrenics do give more idiosyncratic associations to emotional stimuli than do normals. However, as Deering (1963) and Laffal (1955) both stress, the increase in number of unique responses may not be due to affect, but simply because emotional words evoke a larger number of associations than neutral words. In other words, emotional and neutral words should be matched in terms of number of associations they evoke, preferably by choosing words with a small number of associations.

Generally speaking, clinicians make the assumption that unusual themes or perceptual distortions to TAT cards are more revealing than usual themes or perceptual accuracy about personality. Only one study has examined this problem. Miller and Scodel (1955) obtained a usual and an unusual TAT story from each of 35 psychotherapy clients. Six judges then attempted "blind" matching of the 35 common and 35 unusual TAT stories to typescript records of the first two psychotherapy sessions of these patients. While the judges' matching was better than chance, the unusual stories were no better matched than the usual ones. Whether this negative finding provides critical evidence against the hypothesis is open to question, as the findings suffer from the fact that only early psychotherapy data were used. The definition of usual and unusual rested on clinical judgment rather than objective normative data.

At the present time, the ratio ($F + \%$) between accurate and unclear form perception is the best measure of response deviations from the perceptual demands of the Rorschach inkblots. It is not within the scope

of this paper to discuss the problems involved in specifying what constitutes good or poor form, although the solution to them ultimately hinges upon a statistical criterion of response frequency. Hypotheses about the validity of $F + \%$ have stimulated some research, most of which has involved a comparison of clinical groups. Korchin (1960) should be consulted for a brief review of some of these studies. It is sufficient to state that the main evidence concerning $F + \%$ tends to confirm the hypothesis that form accuracy is related to adequacy of perceptual contact with reality. Original research is required, however, in which inkblot stimuli are constructed with known variations in their difficulty level for yielding accurate form percepts. With material conforming to a cumulative homogeneity test model, the resulting threshold measure of contact with test stimuli should correlate significantly with external criterion data. Such a measure should also provide a purer measure of contact with reality than $F + \%$.

Finally, it should be noted that the Tomkins–Horn Picture Arrangement Test (Tomkins and Miner, 1957) rests explicitly upon the assumption that rare and homogeneous responses to similar test stimuli are indicative of significant personality characteristics. While the standardization of this test is excellent, more research is required on the validity of the scoring keys for personality. However, the diagnostic power of this test should ultimately prove to be high.

Summary. The validity of this postulate rests primarily upon form accuracy data of the Rorschach and, to a lesser extent, word association experiments. It is suggested that new material be specifically constructed in order to test its validity and for use in clinical situations.

G. POSTULATE SEVEN: ANXIETY EFFECTS AS A FUNCTION OF STIMULUS GENERALIZATION

With increasing perceived similarity between high-anxiety-evoking situations or figures and their stimulus representations on projective material, anxiety effects should increase by gradual increments.

This postulate rests upon the validity of the phenomenon of stimulus generalization and the assumption that projective techniques will validly reflect anxiety effects. The fact of stimulus generalization seems well established, as does the assumption that anxiety effects are shown by projective instruments. Two recent studies (Walker and Atkinson, 1958; Kenny and Chappell, 1963) have shown that experimentally aroused anxiety is reflected in thematic apperceptive stories. While the ability of the Rorschach to reflect anxiety depends upon type of stress and determinants, Neuringer (1962) concluded, after a review of the literature, that laboratory stress induces lowered productivity, fewer W and P and

more *m* responses. Real life stress leads to more *F* responses and a general reduction in the number of responses, *M*, *m*, and color responses.

Data from three studies are relevant to the present postulate. In connection with a study on the deviant family background of schizophrenics, Baxter and Becker (1962) hypothesized that poor premorbid male schizophrenics come from mother-dominated homes which are high in conflict and aggression and that good premorbid male schizophrenics come from father-dominated homes in which the mother was weak and the father was tyrannical. On the basis of this and related considerations, the authors predicted that poor premorbids should produce more anxiety to a maternal figure than to a paternal figure on the TAT, while good premorbids should manifest the reverse pattern. Cards 6BM (mother-son) and 7BM (father-son) of the TAT were given to 18 poor premorbid schizophrenics and 12 good premorbid schizophrenics. The predicted interaction between anxiety and parental figure was significant, with the poor premorbids showing more anxiety to the mother figure than to the father and the good premorbids showing the opposite pattern. This supports the notion that anxiety is generalized to pictorial representations of the original instigator of anxiety and again reinforces the notion that fruitful predictions must take into account the cue properties of stimulus figures.

Two studies by Epstein and Fenz are relevant to this postulate. In their first study (Epstein and Fenz, 1962), novice parachutists were tested on a word association test with four levels of stimulus relevance to parachuting and a set of general anxiety words. Two parallel forms of the word association test were constructed, since the test was administered on the day of the jump and on a control day. For eight Ss the control day was 2 weeks after a jump and for another eight Ss the control day was 2 weeks before a jump. Sixteen nonparachutists served as controls. In accordance with the notion of stimulus generalization, the results showed that the GSR and the mean reaction time in seconds for the parachutists increased monotonically as a function of stimulus relevance and was steeper on the day of the jump than on the control day. In their other investigation, Fenz and Epstein (1962) employed the same Ss and experimental design. However, two sets of TAT-like stimulus cards were substituted for the word association test. The six pictures in each set contained four unrelated to parachuting, one slightly related, and one strongly related. Congruent with the findings from the word association study, the GSR to the initial impact of the pictures increased as a function of stimulus relevance and was steeper on the day of the jump. Reaction time on the day of the jump was similar between neutral and low-relevance pictures, but was significantly longer on the picture

strongly related to parachuting. Another commonly used indicator of anxiety—frequency of pauses—showed similar results to the reaction time data. It should be pointed out that Epstein and Fenz did not employ the notion of stimulus generalization of fear to predict their results. On the basis of a *modified* version of Miller's (1948, 1959) model of approach-avoidance conflict, they assumed that the gradients of approach and avoidance drives may be added, without regard to sign, to obtain a measure of conflict-instigated activation. Accordingly, they formulated the prediction that the activation gradient will increase as a function of increasing stimulus relevance. With the additional assumption that activation may be measured by autonomic reactivity indicators, such as the GSR, the obtained results also conform to their conflict model. Thus, it would appear that the results of both studies could be reconciled either to a stimulus generalization of fear concept or to a conflict model. In their first study, Epstein and Fenz (1962) argue that activation is shown by the gradient of GSR and not fear, on the grounds that parachutists on both the jump and control day produced larger GSR's to the high-relevance parachuting words than to the anxiety words. This argument is not persuasive on two grounds: (1) The anxiety words were included in the first place solely on the grounds of curiosity, and (2) through prior conditioning, fear may be more attached to high-relevance parachute words (e.g., target, jump, bailout) than to *a priori* chosen anxiety words (e.g., blackout, injury, hurt).

Summary. From these three studies, it would appear that anxiety does generalize as a function of perceived similarity between the original fear-producing situation and the stimulus representation of these events on projective tests.

H. POSTULATE EIGHT: DRIVE-STIMULUS RELEVANCE

For stimuli which have been scaled for a single drive, the following expectations should hold: (a) When only highly structured drive-stimulus material evokes thematic drive content, the individual possesses little or no drive represented by the stimulus material; (b) where minimal, moderate, and strong cues of the drive material evoke drive content, the individual is neurotically motivated and preoccupied with the goal of the drive; (c) when strong-drive Ss are compared with low-drive Ss, the magnitude of thematic drive content differences should gradually increase with increasing stimulus relevance to the drive; and (d) as the number of drive cues increase and the drive has anxiety attached to it, anxiety effects should mount.

In attempting to provide a theoretical discussion of how the fore-going predictions are derived, the author will present a revised version

of a model presented elsewhere (Kenny, 1961). The model makes the following assumptions: (a) Proximal stimuli evoke decoding reactions which are perceptual attempts by the individual to translate or decode proximal stimulation into distal objects; (b) drive stimuli, external or internal in origin, will increase the probability that decoding reactions associated with the drive will occur, that is, drive stimuli narrow selectively the number of decoding reactions that are likely to occur; (c) as the number of decoding reactions instigated by stimuli increases, the arousal level of the person will also increase; (d) individuals seek to maintain moderate levels of arousal, avoiding high or low levels by filtering out stimuli that heighten or lower arousal level; (e) unless anxiety completely blocks the occurrence of decoding reactions related to a drive, anxiety level will interact with internal and external drive cues to increase drive-oriented schematic thought; and (f) drives are assumed to raise the gradient of stimulus generalization.

In terms of these assumptions, various predictions may be made about how individuals will function on projective tests. If a set of projective stimuli has been scaled along a dimension of increasing drive relevance, a random sample of Ss should show a gradual increment in thematic drive content as drive-stimulus relevance increases. Since drives are assumed to raise the entire gradient of stimulus generalization, Ss with high drive should give more thematic drive content to cues further down the dimension of drive relevance than low-drive Ss. Similarly, if a person only gives drive content to highly relevant cues, it may be inferred that drive level is low. On the other hand, if drive content is present to all levels of drive-stimulus relevance, then the inference may be made that drive is excessively high and there is marked preoccupation with the goal of the drive. It should be noted that McClelland et al. (1953) have made similar statements from an inductive base.

The preceding discussion has assumed that the drive state of the organism is not associated with anxiety. What happens to thematic drive content when the drive has anxiety associated with it? If the association of anxiety with drive cues is exceedingly high, there may be a complete blockage of thematic drive content to all stimulus representations of the drive. Since such an event is likely to be rare in permissive testing situations, the question becomes one of predicting what happens to thematic content when anxiety continues to operate. In terms of assumptions (b) and (e), above, highly structured drive stimuli should produce more thematic drive content than any other type of cue and should provide the best discrimination between low- and high-drive Ss. Of course, this prediction is bound to break down in the limiting case when cultural training has produced a strong communality of response to

stimulus representations of drive states. In such a case, most Ss will respond in an identical manner, regardless of drive state.

The implication of the present model that highly structured drive stimuli are the best indicators of thematic drive content is exactly the opposite of that of Epstein and his associates (Epstein and Fenz, 1962; Fenz and Epstein, 1962). Their approach-avoidance conflict model makes the pivotal assumption that approach drives have less steep gradients than avoidance drives as a direct function of increasing stimulus relevance, and, as a consequence, approach or goal-relevant responses should increase to stimuli of low relevance and decrease to stimuli of high relevance. Our survey of the literature will investigate the validity of these two opposing predictions, as well as the other hypotheses.

Several studies have been concerned with the interaction between drive-stimulus relevance and drive level. It is difficult to evaluate the results of these studies because the stimulus properties of the projective stimuli and the precise range of the drive level are not clear. A strongly structured card for one researcher may be a moderately ambiguous card for another. What one investigator may call strong drive, another may call moderate. Until terminological differences between investigators are resolved and a methodology of defining drive-stimulus relevance is standardized, the task of comparing the findings of different studies is almost impossible. In spite of these problems, it does seem possible to draw tentative conclusions from a review of the studies in this area.

1. Drive-Stimulus Relevance and Aggression

Some doll play data, analyzed by Wurtz (1960), may be used to support the prediction that anxiety facilitates the expression of aggression. Making the assumption that more anxiety is aroused if aggression is expressed against parents than other children, Wurtz predicted that more direct and violent aggression will be expressed against parent dolls than child dolls. The doll play results of 150 children showed that the parent dolls were the objects of more violent aggression than child dolls, supporting the general proposition that anxiety interacts with drive to increase drive-oriented schematic thought.

In an attempt to evaluate the influences of maternal permissiveness on aggression, Weatherley (1962) divided female undergraduate students into two groups of 50 in terms of mother's reports as to the degree to which they permitted their children to express aggression as a child. Half of the Ss in each group told TAT stories after they had been aroused to aggression by insulting and depreciating comments, the other half in each group not being exposed to arousal conditions. Two levels of aggressive stimulus relevance were used, with cards 8BM and 18GF

representing high cues and 2 and 6GF of the TAT representing low cues. The analysis of the data yielded a significant triple interaction among the variables of permissiveness, arousal, and stimulus relevance. A breakdown of this interaction effect showed that high maternal-permissive Ss manifested a significant increase in thematic aggression to high aggressive cue cards when they were under arousal conditions Low maternal-permissive Ss did not show this effect. If it is assumed that maternal permissiveness provides better opportunities for the reinforcement of aggression than does low maternal permissiveness, then the finding on the high maternal-permissive Ss confirms the theoretical expectations that highly structured drive stimuli are the best indicators of thematic drive content. Kagan's (1956) study of 118 elementary school children also showed that a picture most suggestive of aggression was more predictive of overt aggression than a less suggestive picture. While the least suggestive picture did not separate the aggressive from the non-aggressive children, more stories of aggressive fighting were produced by the aggressive Ss to the picture most suggestive of fighting than by non-aggressive Ss.

However, other studies on the aggressive drive only provide partial support for our theoretical expectations. Stone (1956) reported that TAT cards with moderate aggressive pull produced the greatest discrimination between assaultive and nonassaultive prisoners. High-aggressive pull pictures were not as discriminatory. In keeping with the present expectations, however, the moderate drive-relevance pictures were more discriminatory than the low-relevance pictures. Since the high-aggressive pictures (13MF, 18GF, and 19BM) have stimulus dimensions that can instigate themes other than aggression, the use of these cards leaves something to be desired. In a study (Hokanson and Gordon, 1958) involving 40 college students, half of the Ss were classified as high expressors of aggression and the other half as low expressors on the basis of a manifest hostility questionnaire. Within each group, half of the Ss were aroused to hostility by listening to a case history of a sadistic juvenile delinquent. Subsequent to arousal, all Ss told stories to a TAT card low in aggression (card 2) and one high in aggression (card 8BM). When the Ss were aroused, the low expressors showed more thematic hostility than the high expressors. With respect to cue relevance, the only significant finding was that the high card produced more thematic aggression than the low card for both groups. Saltz and Epstein (1963) divided male college volunteers into three experiment designs on the basis of self-report measures of hostility, guilt, and conflict over hostility. The three groups were composed of 20 Ss, each of whom scored at the extremes on each of these measures. All Ss were administered six TAT-

type pictures, two of low relevance to aggression, one of high relevance to aggression, and four buffer pictures. In terms of their conflict model, Saltz and Epstein predicted that (a) hostile responses to low-relevance pictures are largely due to drive and hostile responses to high-relevance pictures are due to guilt, and (b) conflict over hostility will cause an overresponse to low-relevance pictures and an underresponse to high-relevance pictures. The data offered equivocal support for their first prediction and no support for the second prediction. More specifically, Saltz and Epstein concluded that the first part of their prediction (a) was confirmed by the findings of a positive relationship between the questionnaire measure of hostility and thematic hostility on the two pictures of low aggressive relevance. Since high relevance was only represented by one picture, the resulting error variance in the thematic score may have prevented any correlation between the high card and self-reported hostility. The second part of their first hypothesis (that guilt is best reflected by high-relevance pictures) was not supported by two of their three analyses of the questionnaire measures.

2. Drive-Stimulus Relevance and Sex

The preceding pattern of results in the Saltz and Epstein study are very similar to those found by Leiman and Epstein (1961) on the sex drive. In this study, 66 unmarried university students were first administered four TAT-like pictures of low-relevance to sex and two pictures of high relevance. They then filled out a questionnaire from which two indices of sexual drive and one index of guilt over sex were derived. One index of sexual drive was based on rate of orgasm, the other on deprivation. The data showed a positive relationship between thematic sex scores and the rate measure; however, no relationship occurred between thematic sexual responses and deprivation. The hypothesis that low-relevance pictures would provide the best measure of sexual drive and high-relevance pictures would be the best measure of guilt was not confirmed. However, a *post hoc* analysis of the individual cards did provide some support for this hypothesis. The other hypothesis of the authors that conflict effects would be shown by an overresponse to low-relevance pictures and an underresponse to high-relevance pictures was not upheld. Such a finding is congruent with the data from the Saltz and Epstein study on aggression.

3. Drive-Stimulus Relevance and Fear

The validity of Epstein's hypotheses about conflict effects, that drive is best measured by pictures of low relevance and that conflict is shown by an overresponse to low-relevance cues and an underresponse to high-

relevance cues, is cast into further doubt by the two studies on parachuting (Fenz and Epstein, 1962; Epstein and Fenz, 1962) previously described. It will be recalled that in both studies novice parachutists were given a word association test and a thematic apperception test on the day of a jump and also on a control day. In the study (Fenz and Epstein, 1962) dealing with thematic pictures, the results showed that parachutists on the day of the jump manifested significantly more thematic *approach* responses to the high-relevance pictures than to the low-relevance pictures. Such a finding casts into doubt the notion that the gradient of avoidance is steeper than that of approach, a pivotal assumption in the conflict model. It should be noted that the obtained finding on the high-relevance picture supports the present author's formulation of stimulus relevance. In the word association study, Epstein and Fenz (1962) found their prediction that parachutists would give more parachute-relevant responses to stimulus words of low relevance and fewer such responses to stimuli of high relevance was not upheld. The parachutists on the day of the jump showed similar approach responses to all levels of stimulus relevance. In terms of their model, Epstein and Fenz also predicted that the number of perceptual misrecognitions of the stimulus words increase as a function of increasing stimulus relevance to parachuting. However, the opposite effect occurred; that is, on the day of the jump the parachutists showed a decrease in misperceptions as stimulus relevance to parachuting increased from low to high. This finding offers strong support for the present author's second assumption. In essence, it follows from this assumption that increases in drive will facilitate veridical perception for stimuli related to the drive.

4. Drive-Stimulus Relevance and Other Drives

Three other published studies also bear on the interaction effect between drive level and stimulus relevance. In examining the effects of about 86 hours' sleep deprivation on TAT stories, Murray (1959) found that the sleep-deprived group showed less thematic sleep content than did the control group and that they had significantly fewer sleep themes than the controls on a high sleep-suggestive card. The two groups did not differ in sleep themes to the low- and medium-sleep card. The experiment raises the general issue of whether sleep deprivation induces a drive. To this author, it seems more plausible to suggest that sleep deprivation will result in a general lowering of a person's arousal level. If this is the case, Murray's results may be expected in terms of our fourth assumption. In an early study on the interaction between the hunger drive and stimulus relevance, Epstein and Smith (1956) found that pictures of low stimulus relevance to hunger were better reflectors of the hunger drive than either

a medium or a high picture-pull card. Unfortunately, only one picture was used at each of the high and moderate picture-pull levels. As already indicated, any findings based upon a single picture are likely to be unreliable.

Murstein (1963a) reported an experiment in which he tested an important part of Atkinson's (1958) theory of human motivation. In terms of this theory it may be predicted that, if the motive to achieve and the motive to avoid failure are held constant, achievement motivation will vary with the probability of success, with the highest possible motivation occurring when the probability is .5. In testing this prediction, Murstein found that a .5 group manifested more over-all achievement imagery than a .1 group; the difference between a .5 and .9 group was not significant. The major finding on stimulus relevance was the fact that pictures of moderate relevance to achievement produced the significant difference between the .5 and .1 group. If it can be safely assumed that the .5 group has a greater need to achieve than the .1 group, then such a finding is not in keeping with the present author's formulation of stimulus relevance. After reviewing the studies concerned with Atkinson's theory, Murstein (1963b) has concluded, however, that the empirical evidence in support of the assumption that a .5 group has the highest achievement motivation is "rather fragile."

Summary. Taken together, do the studies investigating the interaction effect between stimulus relevance and drive support the predictions of the eighth postulate? The investigations of Wurtz (1960), Kagan (1956), and Fenz and Epstein (1962) offer strong support for the model. Most of the other studies offer moderate support for the postulate. There is no support for the Epstein and Fenz hypothesis that conflict will cause an overresponse to low-relevance pictures and an underresponse to high-relevance pictures. Finally, their hypothesis that low-relevance pictures are the best measures of drive is seen to be most suspect when one views it against all the studies which have been reviewed in this section.

I. POSTULATE NINE: AROUSAL LEVEL AS A FUNCTION OF AMBIGUITY

For complex projective material, arousal level varies as a positive function of increasing ambiguity.

It will be recalled that a distinction was made between projective techniques based upon the cumulative homogeneity model and those constructed on the pure relative frequency model. The Rorschach and TAT are the classic illustrations of the latter model. By complex projective material, we mean tests based on the pure relative frequency model. Before discussing the ninth postulate, it will be helpful to discuss the methodological approaches to the concept of ambiguity.

1. Definition of Ambiguity

Broadly speaking, there are two procedural techniques for defining the meaning of ambiguity. One may define a scale of increasing ambiguity in terms of increasing physical impoverishment. Some of the procedures that may be used to induce this kind of ambiguity would be reduction in time of exposure, photographic blurring, random or specific removal of stimuli, incomplete line tracings, reduction in illumination, and the like. Without any wish to imply that impoverishing techniques do not produce interesting experimental results, it is proposed that such operations in no way ensure that ambiguity has been manipulated. To this author, it does not seem to make sense to talk of an ambiguous physical stimulus. A stimulus is only ambiguous if it evokes different perceptual or decoding reactions. Impoverishment of the stimulus does not ensure that there will be an increase in the number of decoding reactions. In fact, their number might actually decrease under some conditions of physical impoverishment.

As this author (Kenny, 1961) proposed elsewhere, an adequate definition of ambiguity must be in terms of decoding reactions. A good quantitative index of ambiguity would take into account the number of alternative decoding reactions to a stimulus and the proportion of cases responding with any given decoding system. This author has proposed that the $\hat{H}$ measure of uncertainty from information theory takes into account these two factors. If we let $p(i)$ indicate the proportion of Ss who give each of the different decoding reactions to a stimulus, then $\hat{H}$ is given by

$$\hat{H} = -\Sigma p(i) \log_2 p(i)$$

For example, if a projective stimulus elicits four different decoding reactions, with the proportion of individuals giving these reactions being .96, .02, .01, and .01, then $\hat{H}$ is equal to .2557. As H increases, perceptual or decoding ambiguity increases. The original article (Kenny, 1961) should be consulted for a more extensive discussion and rationale of the H measure, as might Murstein's (1963b) interesting application of it.

2. Ambiguity and Arousal Level

The assumption is made that increasing ambiguity heightens the arousal level of the individual and that Ss attempt to hold to intermediate levels of arousal by avoiding high arousal levels. The additional operational assumption is made that arousal level can be roughly indicated by ratings on Osgood's (Osgood, Suci, and Tannenbaum, 1957) connotative factors of potency and activity and that high arousal levels will be judged as undesirable, as indicated by low ratings on the evaluative factor. In

other words, ambiguity should correlate positively with potency and activity and negatively with evaluation. This theorizing should be contrasted with the often made assumption that ambiguity facilitates the expression of covert layers of personality. Any golden nuggets (Kenny, 1954, 1961; Kenny and Bijou, 1953; Bijou and Kenny, 1951; Murstein, 1958, 1963b; Gurel and Ullman, 1958; Laskowitz, 1963; Weisskopf, 1950a,b; Weisskopf-Joelson and Lynn, 1953) that such an assumption had in the early days of projective testing have been mined and it is now desirable to move on to a finer-grain analysis of psychological functioning on projective tests. The present postulate attempts such an analysis.

No published research on this postulate has yet appeared. However, this author and his associates (Barbara Long, Lynn Harvey, Atsuko Moriya, and H. Kuechler) have accumulated a large amount of data on the $\hat{H}$ measure and its correlates, which is being prepared for publication. While this is not the place to describe the details of the research, it may be in order to mention that the first phase of the research required various groups of Ss to examine 26 TAT cards for a 20-second period and to describe what they saw. On the basis of these perceptual statements, decoding categories were formed. In order to obtain an index of ambiguity for each card, the uncertainty measure has been applied to this data. The second phase of the research has sought to determine the correlates of the uncertainty measure. Only one finding of this research will be reported, that bearing on the present postulates.

Some unpublished data, collected by Atsuko Moriya under the author's guidance, offer confirmation of the postulate. Male and female university students served as Ss for the derivation of two $\hat{H}$ indices for each of 26 TAT cards, one for each of the sexes. Two other male and female groups rated each of the pictures on 9 bipolar concepts, three concepts for each of the three connotative factors. In order to investigate the predicted relationships, the uncertainty measure rank order of the 26 cards was correlated with each of the three rank orders of the cards based on the three semantic factors. For males, the rank order correlations between ambiguity and evaluation, activity and potency, were —.46, .61, and .58, respectively. For females the corresponding correlations were —.52, .40, and .60, respectively. All six correlations are statistically significant at less than the .05 level. These findings strongly indicate that ambiguity is correlated negatively with the evaluative factor and positively with the activity and potency factors, thereby supporting the postulate that ambiguity is arousal raising. However, a great deal more research is required before the postulate is viewed as anything other than a tentative formulation.

With some evidence in support of the notion that ambiguity is

correlated with approximate indicators of arousal level, the next research question under investigation is, "How does arousal level interact with the existing personality structure within the individual?" An adequate conceptual and empirical answer to this most challenging question would take us far beyond the scope of this paper, which deals with stimulus functions in projective techniques.

III. Postscript

Traditionally the focus of projective techniques has been personality centered, with only a token recognition of the fact that the stimulus might have something to do with behavior. Throughout the review, an attempt has been made to come to grips with the stimulus problem. In order to make an interpretive pathway through a maze of diverse findings, a set of postulates as to how the stimulus articulates with personality functioning is proposed. Lest these postulates and the theorizing back of them cause premature closure, it should be stressed that the present formulation is presented in a tentative and cautious manner. A margin of error in the postulates is expected. It is hoped that they are worded in a sufficiently precise manner to be empirically tested. An extension of the present set of postulates into personality functioning is planned.

Much of the theorizing is not, of course, wholly original. In particular, an attempt has been made to establish a point of contact with recent developments in the perceptual and cognitive areas. Furthermore, it is hoped that the bridge offered between ambiguity and arousal will prove fruitful.

One final statement may be in order. If projective techniques are to survive in psychology, projective personologists must maintain contact with the rapidly expanding knowledge in many areas of psychology, especially perception, cognition, learning, motivation, and measurement theory, and modify their traditional instruments and interpretations in light of new information.

References

Albert, R. S. The role of mass media and the effect of aggressive film content upon children's aggressive responses and identification choices. *Genet. Psychol Monogr.*, 1957, 55, 221-285.

Allen, R. M. The influence of color in the Rorschach test on reaction time in a normal population. *J. proj. Tech.*, 1951, 15, 481-485.

Allen, R. M., Manne, S. H., and Stiff, Margaret P. The role of color in Rorschach's test: a preliminary normative report on a college student population. *J. proj. Tech.*, 1951, 15, 235-242.

Allen, R. M., Manne, S. H., and Stiff, Margaret P. The influence of color on the consistency of responses in the Rorschach test. *J. clin. Psychol.*, 1952, 8, 97-98.

Allen, R. M., Stiff, Margaret P., and Rosenzweig, M. The role of color in Rorschach's test: a preliminary survey of neurotic and psychotic groups. *J. clin. Psychol.*, 1953, 9, 81-83.

Alper, Thelma G. Predicting the direction of selective recall: its relation to ego strength and achievement. *J. abnorm. soc. Psychol.*, 1957, 55, 149-165.

Andrew, G., Walton, R. E., Hartwell, S. W., and Hutt, M. L. The Michigan Picture Test: the stimulus values of the cards. *J. consult. Psychol.*, 1951, 15, 51-54.

Andrew, G., Hartwell, S. W., Hutt, M. L., and Walton, R. E. *The Michigan Picture Test.* Chicago: Science Research Associates, 1953.

Angelini, A. L. Um novo metodo para avaliar a motivacao humano. (A new method of evaluating human motivation.) *Bol. Fac. Filos. Cienc. S. Paulo.*, 1955, No. 207.

Armstrong, M. A. S. Children's responses to animal and human figures in thematic pictures. *J. consult. Psychol.*, 1954, 18, 67-70.

Arnheim, R. Perceptual and aesthetic aspects of the movement response. *J. Pers.*, 1951, 19, 265-281.

Arnheim, R. P. *Art and visual perception.* Berkeley: Univer. of California Press, 1954.

Atkinson, J. W. *Studies in projective measurement of achievement motivation.* Unpublished doctoral dissertation, Univer. of Michigan, Ann Arbor, 1950.

Atkinson, J. W. (Ed.), *Motives in fantasy, action and society.* Princeton, New Jersey: Van Nostrand, 1958.

Auld, F. The influence of social class on tests of personality. *Drew Univer. Studies,* 1952, No. 5, 1-16.

Auld, F., Eron, L. D., and Laffal, J. Application of Guttman's scaling method to the TAT. *Educ. psychol. Measmt.*, 1955, 15, 422-435.

Barnett, I. *The influence of color and shading on the Rorschach test.* Unpublished doctoral dissertation, Univer. of Pittsburgh, Pennsylvania, 1950.

Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.

Barron, F. Threshold for the perception of human movement in inkblots. *J. consult. Psychol.*, 1955, 19, 33-38.

Barron, F. Ego-strength and the management of aggression. In G. S. Welsh and W. G. Dahlstrom (Eds.), *Basic readings in the MMPI in psychology and medicine.* Minneapolis: Univer. of Minnesota Press, 1956, pp. 579-585.

Bartlett, F. C. *Remembering.* London: Cambridge Univer. Press, 1932.

Baughman, E. E. A comparative analysis of Rorschach forms with altered stimulus characteristics. *J. proj. Tech.*, 1954, 18, 151-164.

Baughman, E. E. The role of the stimulus in Rorschach responses. *Psychol. Bull.*, 1958, 55, 121-147.

Baughman, E. E. An experimental analysis of the relationship between stimulus structure and behavior on the Rorschach. *J. proj. Tech.*, 1959, 23, 134-183.

Baxter, J. C., and Becker, J. Anxiety and avoidance behavior in schizophrenics in response to parental figures. *J. abnorm. soc. Psychol.*, 1962, 64, 432-437.

Beck, S. J. Configurational tendencies in Rorschach responses. *Amer. J. Psychol.*, 1933, 45, 433-443.

Belden, A. W., and Baughman, E. E. The effects of figure-ground contrast upon perception as evaluated by a modified Rorschach technique. *J. consult. Psychol.*, 1954, 18, 29-34.

Bellak, L. *The Thematic Apperception Test and the Children's Apperception Test in clinical use.* New York: Grune & Stratton, 1954.

Bellak, L., and Adelman, Crusa. The Children's Apperception Test (CAT). In A. I. Rabin and Mary R. Haworth (Eds.), *Projective techniques with children.* New York: Grune & Stratton, 1960, pp. 62-94.

Berliner, A. The Rorschach determinant in terms of visual psychology. *Optom. Wkly.*, 1955, **20**, 13-17.

Biersdorf, K. R., and Marcuse, F. L. Responses of children to human and to animal pictures. *J. proj. Tech.*, 1953, **17**, 455-459.

Bijou, S. W., and Kenny, D. T. The ambiguity values of TAT cards. *J. consult. Psychol.*, 1951, **15**, 203-209.

Bills, R. E. Animal pictures for obtaining children's projections. *J. clin. Psychol.*, 1950, **6**, 291-293.

Bills, R. E. Self-concepts and Rorschach signs of depression. *J. consult. Psychol.*, 1954, **18**, 135-138.

Bills, R. E., Leiman, C. J., and Thomas, R. W. A study of the validity of the TAT cards and a set of animal pictures. *J. clin. Psychol.*, 1950, **6**, 293-295.

Binder, H. Die Hellunkeldeutungen im psychodiagnostischen experiment von Rorschach. *Schweiz. Arch. Neurol. Psychiat.*, 1933, **30**, 1-67.

Blum, G. S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, **39**, 3-99.

Bohm, E. The Binder chiaroscuro system and its theoretical basis. In Maria A. Rickers-Ovsiankina (Ed), *Rorschach psychology*. New York: Wiley, 1960, pp. 202-222.

Brackbill, G. A. Some effects of color in thematic fantasy. *J. consult. Psychol.*, 1951, **15**, 412-418.

Bradburn, N. M. *N* achievement and father dominance in Turkey. *J. abnorm. soc. Psychol.*, 1963, **67**, 464-468.

Bramel, D. A dissonance theory approach to defensive projection. *J. abnorm. soc. Psychol.*, 1962, **64**, 121-129.

Bramel, D. Selection of a target for defensive projection. *J. abnorm. soc. Psychol.*, 1963, **66**, 318-324.

Brody, G. G. A study of the effects of color on Rorschach responses. *Genet. Psychol. Monogr.*, 1953, **48**, 261-311.

Brosin, H. W., and Fromm, E. O. Rorschach and color blindness. *Rorschach Res. Exch.*, 1940, **4**, 39-70.

Brosin, H. W., and Fromm, E. O. Some principles of gestalt psychology in the Rorschach experiment. *Rorschach Res. Exch.*, 1942, **6**, 1-15.

Bruner, J. S. Perceptual theory and the Rorschach test. *J. Pers.*, 1948, **17**, 157-168.

Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald Press, 1951, pp. 121-147.

Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123-152.

Bruner, J. S., and Postman, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69-77.

Bruner, J. S., and Postman, L. Perception, cognition and behavior. *J. Pers.*, 1949, **18**, 14-31.

Brunswik, E. Distal focusing of perception: size constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, No. 254.

Brunswik, E. *The conceptual framework of psychology*. Chicago, Illinois: Univer. of Chicago Press, 1952. (Int. Encyl. Unified Sci., Vol. I, No. 10.)

Brunswik, E. Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.*, 1955, **62**, 193-217.

Budoff, M. *An investigation of the relative usefulness of animals and persons in a picture-story test for children.* Unpublished master's thesis, Univer. of Chicago, Chicago, Illinois, 1955.

Buker, S. L., and Williams, M. Color as a determinant of responsiveness to Rorschach cards in schizophrenia. *J. consult. Psychol.*, 1951, **15**, 196-202.

Canter, A. *An investigation of the psychological significance of reactions to color on the Rorschach and other tests.* Unpublished doctoral dissertation, Univer. of Iowa, Iowa City, 1951.

Canter, A. The effect of unshaded bright colors in the Rorschach upon the form-color response balance of psychotic patients. *J. proj. Tech.,* 1958, **22,** 390-393.

Carr, A. C. The relation of certain Rorschach variables to expression of affect in the TAT and SCT. *J. proj. Tech.,* 1956, **20,** 137-142.

Child, I. L., Frank, Kitty F., and Storm, T. Self-ratings and TAT: their relations to each other and to childhood background. *J. Pers.,* 1956, **25,** 96-114.

Chodorkoff, B. A note on Bitterman and Kniffen's "Manifest anxiety and perceptual defense." *J. abnorm. soc. Psychol.,* 1955, **50,** 144.

Clark, J. H. Some MMPI correlates of color responses in the group Rorschach. *J. consult. Psychol.,* 1948, **12,** 384-386.

Clark, K. B., and Clark, Mamie P. Racial identification and preferences in Negro children. In T. M. Newcomb and E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947, pp. 169-178.

Cohen-Séat, G., Rebéillard, M. Test filmique thématique. *Rev. int. Filmol.,* 1955, **6,** 111-118.

Cook, R. A. Identification and ego defensiveness in thematic apperception. *J. proj. Tech.,* 1953, **17,** 312-319.

Costello, C. G. The Rorschach records of suicidal patients. *J. proj. Tech.,* 1958, **22,** 272-275.

Cox, B., and Sargent, H. TAT responses of emotionally disturbed and emotionally stable children: clinical judgment versus normative data. *J. proj. Tech.,* 1950, **14,** 61-74.

Cox, F. N., and Sarason, S. B. Test anxiety and Rorschach performance. *J. abnorm. soc. Psychol.,* 1954, **49,** 371-377.

Crumpton, Evelyn. The influence of color on the Rorschach test. *J. proj. Tech.,* 1956, **20,** 150-158.

Deering, G. Affective stimuli and disturbance of thought processes. *J. consult. Psychol.,* 1963, **27,** 338-343.

Douvan, Elizabeth. Social status and success strivings. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society.* Princeton, New Jersey: Van Nostrand, 1958, pp. 509-517.

Dubrovner, R. J., Von Lackum, W. J., and Jost, H. A. A study of the effect of color on productivity and reaction time in the Rorschach test. *J. clin. Psychol.,* 1950, **6,** 331-336.

Edmonston, W. E., and Griffith, R. M. Rorschach content and inkblot structure. *J. proj. Tech.,* 1958, **22,** 394-397.

Epstein, Marilyn. *The reliability and internal consistency of the Thematic Apperception Test.* Unpublished master's thesis, Univer. of British Columbia, Vancouver, Canada, 1964.

Epstein, S., and Fenz, W. D. Theory and experiment on the measurement of approach-avoidance conflict. *J. abnorm. soc. Psychol.,* 1962, **64,** 97-112.

Epstein, S., and Smith, R. Thematic apperception as a measure of the hunger drive. *J. proj. Tech.,* 1956, **20,** 372-384.

Eron, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and nonhospitalized college students. *J. consult. Psychol.,* 1948, **12,** 387-395.

Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.,* 1950, **64,** No. 9 (Whole No. 315).

Exner, J. E. The influence of chromatic and achromatic color in the Rorschach. *J. proj. Tech.*, 1959, 23, 418-425.

Fenz, W. D., and Epstein, S. Measurement of approach-avoidance conflict by a stimulus dimension in a test of thematic apperception. *J. Pers.* 1962, 30, 613-632.

Feshbach, S. The effects of emotional restraint upon the projection of positive affect. *J. Pers.*, 1963, 31, 471-481.

Feshbach, S., and Feshbach, Norma. Influence of the stimulus object upon the complementary and supplementary projection of fear. *J. abnorm. soc. Psychol.*, 1963, 66, 498-502.

Feshbach, S., and Singer, R. D. The effects of fear arousal and suppression of fear upon social perception. *J. abnorm. soc. Psychol.*, 1957, 55, 283-288.

Feshbach, S., Singer, R. D., and Feshbach, Norma. Effects of anger arousal and similarity upon the attribution of hostility to pictorial stimuli. *J. consult. Psychol.*, 1963, 27, 248-252.

Fiske, D. W. Homogeneity and variation in measuring personality. *Amer. Psychologist*, 1963, 18, 643-652.

Forsyth, R. P. The influence of color, shading, and Welsh Anxiety level on Elizur Rorschach content test analyses of anxiety and hostility. *J. proj. Tech.*, 1959, 23, 207-213.

Framo, J. L. *Structural aspects of perceptual development in normal adults: a tachistoscopic study with the Rorschach technique.* Unpublished doctoral dissertation, Univer. of Texas, Austin, 1952.

Friedman, H. Perceptual regression in schizophrenia. An hypothesis suggested by the use of the Rorschach test. *J. proj. Tech.*, 1953, 17, 171-186.

Furuya, K. Responses of school children to human and animal pictures. *J. proj. Tech.*, 1957, 21, 248-252.

Gemelli, A. Le film, procédé d'analyse projective. *Rev. int. Filmol.*, 1951, 2, 135-138.

Gerard, R. U. *Differential effects of colored lights on psychophysiological functions.* Unpublished doctoral dissertation, Univer. of California, Los Angeles, 1958.

Goodman, Mary E. *The genesis of race awareness and attitude.* Unpublished doctoral dissertation, Radcliffe College, Cambridge, Massachusetts, 1946.

Gottesman, I. I. More construct validation of the Ego-Strength scale. *J. consult. Psychol.*, 1959, 23, 342-346.

Grayson, H. M. Rorschach productivity and card preferences as influenced by experimental variation of the color and shading. *J. proj. Tech.*, 1956, 20, 288-296.

Greenbaum, M., Qualtere, T., Carruth, B., and Cruickshank, W. Evaluation of a modification of thematic Apperception Test for use with physically handicapped children. *J. clin. Psychol.*, 1953, 9, 40-44.

Gurel, L., and Ullmann, L. P. Quantitative differences in response to TAT cards: the relationship between transcendence score and number of emotional words. *J. proj. Tech.*, 1958, 22, 432-439.

Guttman, L. The basis for scalogram analysis. In S. A. Stouffer (Ed.), *Studies in social psychology in World War II. Vol. IV. Measurement and prediction.* Princeton, New Jersey: Princeton Univer. Press, 1950, pp. 60-90.

Haase, W. *Rorschach diagnosis, socio-economic class and examiner bias.* Unpublished doctoral dissertation, New York Univer., 1956.

Hartmann, H. Comments on the psychoanalytic theory of the ego. *Psychoanal. stud. Child*, 1950, 5, 74-96.

Hartmann, H. *Ego Psychology and the problem of adaptation.* New York: International Univer. Press, 1958.

Hartwell, S. W., Hutt, M. L., Andrew, Gwen, and Walton, R. E. The Michigan Picture

Test: diagnostic and therapeutic possibilities of a new projective test for children. *Am. J. Orthopsychiat.*, 1951, 21, 124-137.

Haworth, Mary. Films as a group technique. In A. I. Rabin and Mary R. Haworth (Eds.), *Projective techniques with children*. New York: Grune and Stratton, 1960, pp. 177-190.

Hemmindinger, L. Perceptual organization and development as reflected in the structure of Rorschach test responses. *J. proj. Tech.*, 1953, 17, 171-185.

Hemmindinger, L. Developmental theory and the Rorschach method. In Maria Rickers-Ovsiankina (Ed.), *Rorschach psychology*. New York: Wiley, 1960, pp. 58-79.

Henry, W. E. The thematic apperception technique in the study of culture-personality relations. *Genet. Psychol. Monogr.*, 1947, 35, 3-135.

Henry, W. E. *The analysis of fantasy*. New York: Wiley, 1956.

Hochberg, J. E., Triebel, W., and Seaman, G. Color adaptation under conditions of homogeneous visual stimulation (Ganzfeld). *J. exp. Psychol.*, 1951, 41, 153-159.

Hokanson, J. E., and Gordon, J. E. The expression and inhibition of hostility in imaginative and overt behavior. *J. abnorm. soc. Psychol.*, 1958, 57, 327-333.

Holt, R. R. A critical examination of Freud's concept of bound vs. free cathexis. *J. Amer. Psychoanal. Ass.*, 1962, 10, 475-525.

Holtzman, W. H., Iscoe, I., and Calvin, A. D. Rorschach color responses and manifest anxiety in college women. *J. consult. Psychol.*, 1954, 18, 317-324.

Holtzman, W. H., Thorpe, J. S., Swartz, J. D., and Herron, E. W. *Inkblot perception and personality*. Austin: Univer. of Texas Press, 1961.

Horiuchi, Haruyo. A study of perceptual processes of Rorschach cards by tachistoscopic method of movement and shading responses. *J. proj. Tech.*, 1961, 25, 44-53.

Hull, C. L. *Principles of behavior*. New York: Appleton, 1943.

Jensen, A. R. The reliability of projective techniques: review of the literature. *Acta Psychologica*, 1959, 16, 108-136.

Jones, R. A. *Some effects of emotion on the use of words by schizophrenics*. Washington, D. C.: Catholic Univer. American Press, 1957.

Kagan, J. The measurement of overt aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, 52, 390-393.

Kagan, J. Thematic apperceptive techniques with children. In A. I. Rabin and Mary R. Haworth (Eds.), *Projective techniques with children*. New York: Grune and Stratton, 1960, pp. 105-129.

Kelly, E. L. Theory and techniques of assessment. *Annual Review of Psychology*, Vol. 5, Palo Alto: Annual Reviews, 1954, pp. 281-311.

Kenny, D. T. Transcendence indices, extent of personality factors in fantasy responses, and the ambiguity of TAT cards. *J. consult. Psychol.*, 1954, 18, 345-348.

Kenny, D. T. A theoretical and research reappraisal of stimulus factors in the TAT. In J. Kagan and G. S. Lesser (Eds.), *Contemporary issues in thematic apperception methods*. Springfield, Illinois: Charles C Thomas, 1961, pp. 288-310.

Kenny, D. T., and Bijou, S. W. Ambiguity of pictures and extent of personality factors in fantasy responses. *J. consult. Psychol.*, 1953, 17, 283-288.

Kenny, D. T., and Chappell, Marguerite C. Anxiety effects in thematic apperception induced by homogeneous visual stimulation. *J. proj. Tech. and Pers. Assess.*, 1963, 27, 297-301.

Kenny, D. T., Harvey, Lynn, and Wilson, Barbara. Perceptual reactions to the TAT. Unpublished manuscript, 1961.

Kiesler, C., and Singer, R. D. The effects of similarity and guilt on the projection of hostility. *J. clin. Psychol.*, 1963, **19**, 157-162.

Klein, A., and Arnheim, R. Perceptual analysis of a Rorschach card. *J. Pers.*, 1953, **22**, 60-70.

Kohler, F. J., and Steil, A. The use of the Rorschach in involutional melancholia. *J. consult. Psychol.*, 1953, **17**, 365-370.

Korchin, S. J. Form perception and ego functioning. In Maria A. Rickers-Ovsiankina (Ed.), *Projective Psychology.* New York: Wiley, 1960, pp. 109-129.

Korchin, S. J., Mitchell, H. E., and Meltzoff, J. A. A critical evaluation of the Thompson Thematic Apperception Test. *J. proj. Tech.*, 1950, **14**, 445-452.

Kreezer, G. L. A threshold-method for measuring the attention-demand value of stimuli. *Amer. J. Psychol.*, 1958, **71**, 111-122.

Laffal, J. Response faults in word association as a function of response entropy. *J. abnorm. soc. Psychol.*, 1955, **50**, 265-270.

Lasaga y Travieso, J. E., and Martinez-Arango, C. Some suggestions concerning the administration and interpretation of the TAT. *J. Psychol.*, 1946, **22**, 117-163.

Laskowitz, D. Degree of pictorial ambiguity and fantasy evocation: an appraisal of the nonmonotonic hypothesis. *Perceptual and Motor Skills*, 1963, **16**, 187-193.

Lazarus, R. S. An experimental analysis of the influence of color on the protocol of the Rorschach test. *J. Pers.*, 1948, **17**, 182-185.

Lazarus, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, **44**, 506-516.

Lazarus, R. S., and Oldfield, Margaret. Rorschach responses and the influence of color. *J. Pers.*, 1955, **23**, 356-372.

Leiman, A. H., and Epstein, S. Thematic sexual responses as related the sexual drive and guilt. *J. abnorm. soc. Psychol.*, 1961, **63**, 169-175.

Lennep, D. J. van. *Four-pictures test.* The Hague: Martinus Nijhoff, 1948.

Lerner, E., and Murphy, L. B. Methods for the study of personality in young children. *Monogr. Soc. Res. Child Developm.*, 1941, **6**, No. 4.

Lesser, G. S. Application of Guttman's scaling method to aggressive fantasy in children. *Educ. psychol. Measmt.*, 1958, **18**, 543-552.

Lesser, G. S., Krawitz, Rhode N., and Packard, Rita. Experimental arousal of achievement motivation in adolescent girls. *J. abnorm. soc. Psychol.*, 1963, **66**, 59-66.

Levine, R., Chein, I., and Murphy, G. The relation of intensity of a need to the amount of perceptual distortion: A preliminary report. *J. Psychol.*, 1942, **13**, 283-293.

Levitt, E. E. Alleged Rorschach indices in children. *J. proj. Tech.*, 1957, **21**, 261-264.

Levitt, E. E., and Grosz, H. D. A comparison of quantifiable Rorschach anxiety indicators in hypnotically induced anxiety and normal states. *J. consult. Psychol.*, 1960, **24**, 31-34.

Levy, D. M. Studies in sibling rivalry. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1937, No. 2.

Levy, E. Stimulus value of Rorschach cards for children. *J. proj. Tech.*, 1958, **22**, 293-296.

Light, B. H. A further test of the Thompson TAT rationale. *J. abnorm. soc. Psychol.*, 1955, **51**, 148-150.

Lindzey, G. Thematic Apperception Test: interpretive assumptions and relative empirical evidence. *Psychol. Bull.*, 1952, **49**, 1-25.

Lindzey, G. *Projective techniques and cross-cultural research.* New York: Appleton, 1961.

Lindzey, G., and Goldberg, M. Motivational differences between male and female as measured by the Thematic Apperception Test. *J. Pers.*, 1953, **22**, 101-117.

Lindzey, G., & Herman, P. S. Thematic Apperception Test: a note on reliability and situational validity. *J. proj. Tech.*, 1955, **18**, 36-42.

Lindzey, G., and Kalnins, D. Thematic Apperception Test: some evidence bearing on the "hero assumption." *J. abnorm. soc. Psychol.*, 1958, **57**, 76-83.

Little, K. B. Connotations of the Rorschach inkblots. *J. Pers.*, 1959, **27**, 397-406.

Loiselle, R. H., and Kleinschmidt, Ann. A comparison of the stimulus value of Rorschach inkblots and their percepts. *J. proj. Tech.*, 1963, **27**, 191-194.

Lubin, N. M. The effect of color in the TAT on productions of mentally retarded subjects. *Amer. J. ment. Defic.*, 1955, **60**, 366-370.

Lubin, N. M., and Wilson, M. O. Picture test identification as a function of "reality" (color) and similarity of picture to subject. *J. gen. Psychol.*, 1956, **54**, 31-38.

McArthur, C. Personality differences between middle and upper classes. *J. abnorm. soc. Psychol.*, 1955, **50**, 247-254.

McClelland, D. C. Methods of measuring human motivation. In J. W. Atkinson (Ed.). *Motives in fantasy, action, and society.* Princeton, New Jersey: Van Nostrand, 1958, pp. 7-42.

McClelland, D. C., Atkinson, J. W., Clark, R. A., and Lowell, E. I. *The achievement motive.* New York: Appleton, 1953.

Maccoby, Eleanor E., and Wilson, W. C. Identification and observational learning from films. *J. abnorm. soc. Psychol.*, 1957, **55**, 76-87.

McIntyre, C. J. Sex, age, and iconicity as factors in projective film tests. *J. consult. Psychol.*, 1954, **18**, 475-477.

Mainord, F. R., and Marcuse, F. L. Responses of disturbed children to human and to animal pictures. *J. proj. Tech.*, 1954, **18**, 475-477.

Mann, L. The relation of Rorschach indices of extratension and introversion to a measure of responsiveness to the immediate environment. *J. consult. Psychol.*, 1956, **20**, 114-118.

Mayer, J., and Binz, E. Stimulus values of Rorschach cards. *J. consult. Psychol.*, 1961, **17**, 186-187.

Meer, B., and Singer, J. L. A note on the "father" and "mother" cards in the Rorschach inkblots. *J. consult. Psychol.*, 1950, **14**, 482-484.

Meyer, B. T. An investigation of color chock in the Rorschach test. *J. clin. Psychol.*, 1951, **7**, 267-270.

Miller, J. S., and Scodel, A. The diagnostic significance of usual and unusual TAT stories. *J. consult. Psychol.*, 1955, **19**, 91-95.

Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.

Miller, N. E. Liberalization of basic S-R concepts: extensions to conflict behavior, motivation and social learning. In S. Koch. (Ed.), *Psychology: a study of a science.* New York: McGraw-Hill, 1959, Vol. 2.

Moore, Mary E., and Schwartz, M. U. The effect of the sex of the frustrated figure on responses to the Rosenzweig P-F Study. *J. proj. Tech.*, 1963, **27**, 195-199.

Morgan, C. D., and Murray, H. A. A method for investigating fantasies: the Thematic Apperception Test. *Arch. Neurol. & Psychiat.*, 1935, **34**, 289-306.

Moriya, Atsuko. *The effects of culture on ambiguity and connotative meanings of Thematic Apperception Cards.* Unpublished master's thesis, Univer. of British Columbia, Vancouver, Canada, 1962.

Murray, E. J. Conflict and repression during sleep deprivation. *J. abnorm. soc. Psychol.*, 1959, **59**, 95-101.

Murray, H. A. The effect of fear on the estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, 4, 310-329.

Murray, H. A. *Thematic Apperception Test*. Cambridge, Massachusetts: Harvard Univer. Press, 1943.

Murstein, B. I. The relationship of stimulus ambiguity on the TAT to productivity of themes. *J. consult. Psychol.*, 1958, 22, 348.

Murstein, B. I. The role of the stimulus in the manifestation of fantasy. In J. Kagan and G. S. Lesser (Eds.), *Contemporary issues in thematic apperceptive methods.* Springfield, Illinois: Charles C Thomas, 1961, pp. 229-273.

Murstein, B. I. The relationship of expectancy of reward to achievement performance on an arithmetic and thematic test. *J. consult. Psychol.*, 1963, 27, 394-399. (a)

Murstein, B. I. *Theories and research in projective techniques:* Emphasizing the TAT. New York: Wiley, 1963. (b)

Neuringer, C. Manifestations of anxiety on the Rorschach test. *J. proj. Tech.*, 1962, 26, 318-326.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. *The measurement of meaning.* Urbana: Univer. of Illinois Press, 1957.

Perlman, Janet A. Color and the validity of the Rorschach 8-9-10 per cent. *J. consult. Psychol.*, 1951, 15, 122-126.

Phillipson, H. *The object relations technique.* Chicago, Illinois: The Free Press, 1955.

Piaget, J. *Psychology of intelligence.* London: Routledge and Kegan Paul, 1950.

Piotrowski, Z. A. The Thematic Apperception Test of a schizophrenic interpreted according to new rules. *Psychoanal. Rev.*, 1952, 39, 230-251.

Postman, L., Bruner, J. S., and McGinnies, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, 83, 148-153.

Postman, L., and Solomon, R. L. Perceptual sensitivity to completed and uncompleted tasks. *J. Pers.*, 1949, 18, 347-357.

Purcell, K. The TAT and antisocial behavior. *J. consult. Psychol.*, 1956, 20, 449-456.

Rabin, A. I. A contribution to the "meaning" of Rorschach's inkblots via the semantic differential. *J. consult. Psychol.*, 1959, 23, 368-372.

Rickers-Ovsiankina, Maria. Some theoretical considerations regarding the Rorschach method. *Rorschach Res. Exch.*, 1943, 7, 41-53.

Rickers-Ovsiankina, Maria. A synopsis of psychological premises underlying the Rorschach. In Maria A. Rickers-Ovsiankina (Ed.), *Rorschach psychology.* New York: Wiley, 1960, pp. 3-22.

Riess, B. F., Schwartz, E. K., and Cottingham, A. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, 45, 700-709.

Riessman, F., and Miller, S. M. Social class and projective tests. *J. proj. Tech.*, 1958, 22, 432-439.

Rorschach, H. *Psychodiagnostics: a diagnostic test based on perception.* Bern, Switzerland: Huber, 1942 (1st ed. 1921).

Rosen, E. Symbolic meanings in the Rorschach cards: a statistical study. *J. clin. Psychol.*, 1951, 7, 239-244.

Rosen, E. Connotative meanings of Rorschach inkblots, responses and determinants. *J. Pers.*, 1960, 28, 413-426.

Rosenzweig, S. Idiodynamics in personality theory with special reference to projective methods. *Psychol. Rev.*, 1951, 58, 213-223.

Rosenzweig, S., and Fleming, Edith. Apperceptive norms for the Thematic Apperception Test: II. An empirical investigation. *J. Pers.*, 1949, 17, 483-503.

Rust, R. U. Some correlates of the movement response. *J. Pers.*, 1948, 16, 369-401.

Saltz, G., and Epstein, S. Thematic hostility and guilt responses as related to self-reported hostility, guilt, and conflict. *J. abnorm. soc. Psychol.*, 1963, **67**, 469-479.

Sanford, R. N. The effects of abstinence from food upon imaginal processes: A preliminary experiment. *J. Psychol.*, 1936, **2**, 145-159.

Sanford, R. N. The effects of abstinence from food upon imaginal processes: a further experiment. *J. Psychol.*, 1937, **3**, 145-149.

Sappenfield, B. R. Perception of masculinity-femininity in Rorschach blots and responses. *J. consult. Psychol.*, 1961, **17**, 373-376.

Sappenfield, B. R., and Buker, S. L. Validity of the Rorschach 8-9-10 per cent as an indicator of responsiveness to color. *J. consult. Psychol.*, 1949, **13**, 268-271.

Schachtel, E. On color and affect. *Psychiatry*, 1943, **6**, 393-409.

Schleifer, M. J., and Hire, A. W. Stimulus value of Rorschach inkblots expressed as trait and affective characteristics. *J. proj. Tech.*, 1960, **24**, 164-170.

Schwartz, E. K., Riess, B. F., and Cottingham, Alice. Further critical evaluation of the Negro version of the TAT. *J. proj. Tech.*, 1951, **15**, 394-400.

Seeman, W., and Marks, P. A. A study of some "test dimensions" conceptions. *J. proj. Tech.*, 1962, **26**, 469-473.

Shapiro, D. A perceptual understanding of color response. In Maria A. Rickers-Ovsiankina (Ed.), *Rorschach Psychology*. New York: Wiley, 1960, pp. 154-201.

Sherif, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.

Shneidman, E. S. Schizophrenia and the MAPS test. *Genet. Psychol. Monogr.*, 1948, **38**, 145-224.

Siipola, Elsa M. The influence of color on reactions to inkblots. *J. Pers.*, 1950, **18**, 358-382.

Siipola, Elsa M., Kuhns, Florence, and Taylor, V. Measurement of the individual's reactions to color in inkblots. *J. Pers.*, 1950, **19**, 153-171.

Silverstein, A. B. Identification with same-sex and opposite-sex figures in thematic apperception. *J. proj. Tech.*, 1959, **23**, 73-75.

Singer, J. L., and Spohn, H. E. Some behavioral correlates of Rorschach's experiencer type. *J. consult. Psychol.*, 1954, **18**, 1-9.

Singer, R. D. A cognitive view of rationalized projection. *J. proj. Tech.*, 1963, **27**, 235-243.

Singer, R. D., and Feshbach, S. Effects of anxiety arousal in psychotics and normals upon the perception of anxiety in others. *J. Pers.*, 1962, **30**, 574-587.

Solomon, R. L., and Howes, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.

Spivak, G., Levine, M., and Sprigle, H. Barron M threshold values in emotionally disturbed adolescents. *J. proj. Tech.*, 1958, **22**, 446-449.

Stein, M. I. Personality factors involved in the temporal development of Rorschach responses. *J. proj. Tech.*, 1949, **13**, 355-414.

Stein, M. I. Clinical psychology and the propaedeutic science. *J. proj. Tech.*, 1951, **15**, 401-404.

Sterling, M. E. *Color shock on the Rorschach test*. Unpublished doctoral dissertation, Univer. of Kentucky, Lexington, 1950.

Stone, H. The TAT aggressive content scale. *J. proj. Tech.*, 1956, **20**, 445-452.

Stone, L. J., and Fiedler, Mariam F. The Rorschachs of selected groups of children in comparison with published norms: II. The effect of socio-economic status on Rorschach performance. *J. proj. Tech.*, 1956, **20**, 276-279.

Stricker, G. Stimulus properties of the Blacky Pictures Test. *J. proj. Tech.*, 1963, **27**, 244-247.

Swartz, M. B. The role of color in influencing responses to the Rorschach test: an experimental investigation of the validity of the color shock hypothesis as a sign of neurotic disturbance and as a phenomenon induced by the color stimulus. Unpublished doctoral dissertation, New York Univer., New York, 1953.

Symonds, P. M. *Adolescent fantasy.* New York: Columbia Univer. Press, 1949.

Taniguchi, M., De Vos, G., and Murakami, E. Identification of mother and father cards on the Rorschach by Japanese normal and delinquent adolescents. *J. proj. Tech.*, 1958, **22**, 453-460.

Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.

Thompson, C. E. The Thompson modification of the Thematic Apperception Test. *Rorschach Res. Exch.*, 1949, **13**, 469-478.

Thompson, C. E., and Bachrach, J. The use of color in the Thematic Apperception Test. *J. proj. Tech.*, 1951, **15**, 173-184.

Tolar, A. The "meaning" of the Bender-Gestalt test designs: A study in the use of the semantic differential. *J. proj. Tech.*, 1960, **24**, 433-438.

Tomkins, S. S., and Miner, J. B. *The Tomkins-Horn pictures arrangement test.* New York: Springer, 1957.

Underwood, B. J. Stimulus selection in verbal learning. In C. N. Cofer and Barbara S. Musgrove (Eds.), *Verbal behavior and learning: problems and processes.* New York: McGraw-Hill, 1963, pp. 33-48.

Veroff, J. *A projective measure of achievement motivation in adolescent males and females.* Unpublished honors thesis, Wesleyan Univer., Middletown, Connecticut, 1950.

Veroff, J. Thematic apperception in a nationwide sample survey. In J. Kagan and G. S. Lesser (Eds.), *Contemporary issues in thematic apperceptive methods.* Springfield, Illinois: Charles C Thomas, 1961, pp. 83-111.

Veroff, J., Wilcox, S., and Atkinson, J. W. The achievement motive in high school and college age women. *J. abnorm. soc. Psychol.*, 1953, **48**, 108-119.

Veroff, J., Atkinson, J. W., Feld, Sheila C., and Guria, G. The use of thematic apperception to assess motivation in a nationwide interview study. *Psychol. Monogr.*, 1960, **74** (Whole No. 499).

Walker, E. L., and Atkinson, J. W. The expression of fear-relation motivation in thematic apperception as a function of proximity to an atomic explosion. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* New York: Van Nostrand, 1958, pp. 143-159.

Wallen, R. The nature of color shock. *J. abnorm. soc. Psychol.*, 1948, **43**, 346-356.

Weatherley, D. Maternal permissiveness toward aggression and subsequent TAT aggression. *J. abnorm. soc. Psychol.*, 1962, **65**, 1-5.

Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390. (a)

Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.*, 1950, **29**, 407-416. (b)

Weisskopf, E. A., and Dunlevy, G. P. Bodily similarity between subjects and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1952, **47**, 441-445.

Weisskopf-Joelson, Edith A., and Foster, Helen. An experimental study of the effect of stimulus narration upon projection. *J. proj. Tech.*, 1962, **26**, 366-370.

Weisskopf-Joelson, E. A., and Lynn, D. B. The effect of variations in ambiguity on projection in the Children's Apperception Test. *J. consult. Psychol.*, 1953, 17, 67-70.

Weisskopf-Joelson, E. A., and Money, L. Facial similarity between subject and central figure in the TAT as as influence on projection. *J. abnorm. soc. Psychol.*, 1953, 48, 341-344.

Welsh, G. S. Factor dimensions A and R. In O. S. Welsh and W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. of Minnesota Press, 1956, pp. 264-281.

Werner, H. The concept of development from a comparative and organismic point of view. In D. Harris (Ed.), *The content of development*. Minneapolis: Univer. of Minnesota Press, 1957, pp. 125-148.

Wertheimer, M. Perception and the Rorschach. *J. proj. Tech.*, 1957, 21, 209-216.

Wexner, Lois B. The degree to which colors (hues) are associated with mood tones. *J. appl. Psychol.*, 1954, 38, 432-435.

Williams, H. L., and Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, 17, 261-264.

Williams, H. L., and Lawrence, J. F. Comparison of the Rorschach and MMPI by means of factor analysis. *J. consult. Psychol.*, 1954, 18, 193-197.

Winne, J. F. A peak of neuroticism. An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, 7, 117-122.

Wittenborn, J. R. Some Thematic Apperception Test norms and a note on the use of the test cards in the guidance of college students. *J. clin. Psychol.*, 1949, 5, 157-161.

Wittenborn, J. R. The implications of certain assumptions involved in the use of the Thematic Apperception Test. *J. consult. Psychol.*, 1950, 14, 216-225. (a)

Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, 14, 261-267. (b)

Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950, 14, 469-472. (c)

Wittenborn, J. R., and Holzberg, J. D. The Rorschach and description diagnosis. *J. consult. Psychol.*, 1951, 15, 460-463.

Woodworth, R. S. *Dynamics of behavior*. New York: Holt, 1958.

Wurtz, K. R. Some theory and data concerning the attenuation of aggression. *J. abnorm. soc. Psychol.*, 1960, 60, 134-136.

York, R. H. *The effect of color in the Rorschach test and in selected intellectual tasks.* Unpublished doctoral dissertation, Boston Univer., Boston, Massachusetts, 1951.

Zax, M., and Benham, F. G. The stimulus value of the Rorschach inkblots as perceived by children. *J. proj. Tech.*, 1961, 25, 233-237.

Zax, M., and Loiselle, R. H. The influence of card order on the stimulus value of the Rorschach inkblots. *J. proj. Tech.*, 1960, 24, 218-221. (a)

Zax, M., and Loiselle, R. H. Stimulus value of Rorschach inkblots as measured by the Semantic Differential. *J. clin. Psychol.*, 1960, 16, 160-163. (b)

Zax, M., Loiselle, R. H., and Karras, S. Stimulus characteristics of Rorschach inkblots as perceived by a schizophrenic sample. *J. proj. Tech.*, 1960, 25, 439-443.

Zulliger, H. *Einführung in den Behn-Rorschach test*. Bern, Switzerland: Huber, 1946.

Zulliger, H. *Der Z-Test: ein formdent-Verfahren zur psychologischen Untersuchung von Gruppen*. Bern, Switzerland: Huber, 1948.

AUTHOR INDEX

Numbers in italics indicate the pages on which the complete references are listed.

SUBJECT INDEX

A

Acquiescence set, 234
Alcoholism and repression-sensitization, 190-191
Altrocchi Expressor Index, 203
Anxiety
  awareness of, 180-181
  conditioning in, 121-128
  Rorschach chiaroscuro, 328-330
  stimulus generalization, 332-334
  Welsh Anxiety Scale, 192
  (see also *Manifest Anxiety Scale*)
Arousal
  ambiguous stimuli, 340-343
  conditioning in, 124-128
  schizophrenia, 26-27, 37-40
Atomic war, belief in likelihood of, 199
Attention
  metamphetamine, 10-11
  narrowing in stress, 11-12
    with incentives, 13
  reticular formation, 12-13
  schizophrenia, 20-21, 26-27
  size constancy, 9-10
Authoritarianism,—dogmatism, 199

B

Barron's Achromatic Inkblots, 298
Bias
  communication of, 106-110
  early data returns, 105-106
  experimenter, 70-114
  experimenter attributes and, 84, 99-100
  experimenter expectancy and, 87-97
  experimenter modeling, 85-87
  intentional effects (cheating), 82-84
  observer effects, 80-82
  response sets, 234
  unusual incentives, 97-98
Bills Index of Adjustment and Values, 187
Buss Durkee Inventory, 189

C

Canadian-Japanese, TAT compared, 304-305

Chapin's Social Insight Test, 223
Cheating, in experiments, 82-84
"Clever Hans," 89-90
Conditionability, and personality, 140-144
Conditioning
  alpha blocking, 154-156
  anxiety and, 121-128
  eye-blink, 144-145
  GSR, 145-150
  introversion-extroversion, 118-121
  personality, and, 155-168
  psychosis, 130-132
  Russian personality studies, 133-136
  salivary, 150-154
  set, 138-141
  therapies, 161-163
Couch-Keniston Scale, 199
C.P.I., 194-196
Criminals, repression-sensitization, 197
Cronbach's component, interpersonal perception, 227-230
Crowne-Marlowe Social Desirability Scale, 198-199

D

Deviation Hypothesis, 191-192
Dogmatism Scale, Rokeach, 199

E

Edwards Social Desirability Scale, 197, 198, 209, 210
Empathy, morale and, 230
Experimental neurosis, 156-161

F

F-Scale, 199
Forgetting, in repression, 178

G

Gough's Opinion Prediction Scale, 223
Grigg-Thorpe Adjective Check List, 191-192
Guilford's Inventory of Factors, 200

H

Hereford Parent Attitude Survey, 201
Holtzman Inkblot Test, 189